Dynamic Programming Deconstructed[1]

Qingyin Ma and John Stachurski

Research School of Economics, Australian National University

July 3, 2018

Abstract. Dynamic programming is a foundational technique for modeling economic activity. Part of the art of dynamic programming is to manipulate the Bellman equation into a relatively advantageous form without breaking its link to optimality. In this paper we provide a theoretical framework and a set of results that transform this art into a science. These results (a) clarify the link between existing manipulations from economic applications and optimality of the resulting policies, (b) elucidate the possible extent of such manipulations—and when this link breaks down, (c) clarify the connection between contractivity of the modified Bellman operators and Bellman's principle of optimality, (d) use manipulations of the Bellman equation to extend the set of algorithms for obtaining optimal policies, (e) establish new applications of these manipulations, such as tackling problems with unbounded rewards, (f) extend the set of models that can be transformed to include recursive preferences and other forms of nonseparability, and (g) use these methods to simplify the Bellman equation in a range of recent applications.

JEL Classifications: C61, E00

Keywords: Dynamic programming, optimality

1. Introduction

Dynamic programming is central to the analysis of intertemporal planning problems in economics, finance and operations research.[2] When combined with statistical learning, dynamic programming also drives a number of strikingly powerful algorithms in

[1]We thank Jaap Abbring, Boyan Jovanovic, Fedor Iskhakov, Takashi Kamihigashi, Larry Liu and Daisuke Oyama for valuable feedback and suggestions. Financial support from ARC Discovery Grant DP120100321 is gratefully acknowledged.

Email addresses: `qingyin.ma@anu.edu.au`, `john.stachurski@anu.edu.au`

[2]For surveys and textbook treatments of intertemporal problems that include applications of dynamic programming across various subfields, see, for example, Rust (1996a), Duffie (2010), Pissarides (2000), Bergemann and Valimaki (2017), Bertsekas (2017) or Ljungqvist and Sargent (2018).

artificial intelligence and automated decision systems.[3] Its effectiveness lies in a form of dimensionality reduction, which in turn rests on recursive structure generated by time's unidirectional motion.

Dynamic programming is most efficient when the problem in question is low dimensional, when the associated policy and value functions are defined on relatively small finite sets or are smooth and easily approximated, and when rewards are bounded, so that the standard contraction mapping arguments apply.[4] Conversely, dynamic programming is hampered when the state space is high dimensional, when the value function contains jumps or other irregularities, or when rewards are unbounded.[5]

One way to mitigate these problems is to change the angle of attack by rearranging the planning problem. The idea is to *preserve the optimal policy while modifying the Bellman equation into a more advantageous form.* Such manipulations are henceforth called *plan factorizations* for reasons described below. An important early example is Jovanovic (1982), in which a firm decision problem is solved using a modified Bellman equation based on an action-contingent, time-shifted version of the value function. In particular, the value of a firm is computed from the point in time where the decision to continue has already been made, and firms act optimally in all subsequent periods. The resulting functional equation turns out to be of lower dimension than the standard Bellman equation (which assumes that firms act optimally at the current stage as well).

Another early example of plan factorization can be found in the foundational study of regenerative optimal stopping by Rust (1987), where a modified Bellman equation is constructed from expected value, conditional on the decision to either stop or continue in the current period and behave optimally thereafter. The objective is to simplify the Bellman equation and hence the estimation of underlying parameters. Expected value methods are now widely applied in the fields of discrete choice and structural dynamics to simplify estimation or reduce computational complexity.[6]

These and other plan factorizations have since been adopted across a huge array of economic applications. For example, in the context of dynamic discrete choice

[3]For an overview see Kochenderfer (2015).

[4]Rust (1996b) provides a valuable discussion of practical aspects of dynamic programming.

[5]Additional difficulties arise when preferences are not additively separable over time, a point we return to below.

[6]The set of applications is too large to list but relatively theoretical treatments can be found in Aguirregabiria and Mira (2002) and Norets (2010).

models, Kristensen et al. (2018) consider three versions of the Bellman equation, corresponding to the value function, expected value function and "integrated value function" respectively. Winberry (2018) tackles a household problem using a plan factorization that integrates transient shocks out of the value function in order to act on a lower dimensional space than the standard value function. Further recent applications of dynamic programming that involve substantial manipulations of the Bellman equation can be found in Fajgelbaum et al. (2017), Schaal (2017), Abbring et al. (2018), Bloom et al. (2018) and many other papers.

When viewed as a whole, these contributions raise many questions on the set of possible plan factorizations in dynamic programming and the relationship between them. For example, if we take any one of these plan factorizations in any given context, is it *always* the case that this transformation leaves the optimal policy invariant, in the sense that the policy computed from the fixed point of the transformed Bellman operator is equal to the optimal policy? This is nontrivial because the transformations in question are not in general bijective—indeed they are often constructed precisely so that their range space is lower dimensional than their domain.

More generally, given a fixed dynamic programming problem, do the Bellman operators corresponding to all possible plan factorizations have unique fixed points? Does existence and uniqueness of a fixed point in one case imply existence and uniqueness for the others? Does convergence of the successive iterates of any one of these operators imply convergence of the others? If one of these operators is a contraction mapping, does this confer some form of optimality on the associated policy in the same way that it does for the standard Bellman operator?

In addition, the relative rate of convergence for the respective Bellman operators is also unknown. Assuming that all operators have unique fixed points that are meaningful in terms of optimality, is it the case that they all converge at the same rate? Could it be that one of these operators converges to a given level of tolerance (or produces a policy of higher value) with a smaller number of steps? Might it be the case that one sequence of iterates always converges while some of the others fail to converge at all?

There are additional questions that concern algorithms. Under the standard formulation of the dynamic programming problem, we have access to methods beyond value function iteration that can be effective in certain settings, such as policy iteration and modified policy iteration. Do we always have analogous methods under any given plan factorization? Do they have the same convergence properties? How do we deploy them?

An extra layer of complication arises if we wish to address these questions while at the same time accommodating the kinds of modifications to intertemporal preferences increasingly used in economic applications. For example, how do the answers to the questions posed above change if, say, the additively separable preferences that lie behind the plan factorization used by Winberry (2018) are replaced by Epstein–Zin preferences (see, e.g., Kaplan and Violante (2017) or Schorfheide et al. (2018)), or if we introduce a desire for robustness, or intertemporal risk sensitivity (Hansen and Sargent (2008))?[7]

Addressing all of the questions listed above has obvious value, and yet it is perhaps even more valuable to consider whether or not there are significant new applications of plan factorizations that have not yet been realized, and that might be facilitated or brought to light by a systematic treatment. For example, while the manipulations discussed in the applications listed above have been adopted either to reduce dimension, facilitate estimation or improve intuition, it is natural to ask whether similar manipulations can help with other stumbling blocks for dynamic programming encountered in applications, such as unbounded reward functions or failure of contractivity.

In this paper we construct a theory of plan factorizations in dynamic programming that can be used to address all of the questions listed above. The analysis is general enough to include all plan factorizations used to date as special cases. In addition, the theory is embedded in an abstract dynamic programming framework that allows for both standard and nonstandard preferences, including Epstein–Zin preferences, ambiguity aversion, desire for robustness and risk sensitivity.

One theoretical contribution of the paper is to provide conditions under which all of the possible alternative timings have "equal rights," in the sense that, when the modified value function associated with a given plan factorization satisfies the similarly modified Bellman equation, Bellman's principle of optimality applies: at least one optimal policy exists, and the set of optimal policies is characterized by the pointwise maximality of their actions vis-a-vis the modified Bellman equation. Under the same conditions we show that contractivity and other forms of asymptotic stability for the modified Bellman operators associated with these plan factorizations lead to

[7]Further discussion of nonseparable preferences can be found in Epstein and Zin (1989), Marinacci and Montrucchio (2010), Bloise and Vailakis (2018), Ju and Miao (2012), Hansen et al. (2006), Tallarini (2000), Wang et al. (2016), Bäuerle and Jaśkiewicz (2018) and many other sources.

existence of optimal policies and methods of computing these policies that parallel the traditional methods. The conditions are typically satisfied in applications and center around monotonicity with respect to future value.

We also provide examples showing that the monotonicity requirements in these results cannot be dropped and discuss the problems that can result from failure of monotonicity. While such problems are unlikely to be encountered with additively separable preferences and standard plan factorizations, they do need to be considered when we switch to more sophisticated preferences or adopt less standard plan factorizations.

Furthermore, we show that, with or without monotonicity, the successive iterates of these modified Bellman operators, of which there exist an infinite number due to the infinity of possible plan factorizations, all converge at the same rate in one sense: the n-th iterate of the Bellman operator can alternatively be produced by iterating $n-1$ times with a modified Bellman operator and then performing at most two manipulations (and vice versa). At the same time, if we think of convergence in terms of a specific metric over functions, we find it possible that one operator converges at a geometric rate while another fails to converge at all.

We also treat algorithms for computing optimal policies in addition to successive iteration methods. We focus on so-called modified policy iteration, which contains policy iteration and value function iteration as special cases, and which can typically be tuned to converge faster than either one in specific applications. We show that, under a combination of contractivity and the same monotonicity conditions mentioned above, the sequence of policies generated by a refactored version of modified policy iteration converge to optimality when measured in terms of the lifetime value they produce.

Regarding applications, plan factorizations have most often been adopted to reduce dimensionality, exploit the smoothing properties of conditional expectations or provide a more suitable setting for some form of estimation. We generalize and extend these ideas, applying them in a range of new settings. To illustrate our theoretical results, we use them to simplify the decision problems presented in a range of recent work, with applications in pairwise difference estimation, optimal savings with stochastic discounting, targeted government transfers, consumer bankruptcy, optimal default and portfolio choice with annuities.

In addition, we provide a new application of plan factorizations, showing how a certain transformation can map certain dynamic programs with unbounded rewards into bounded problems, where standard contraction methods based on supremum norms can be applied. To illustrate the value of this approach, we use a plan factorization to convert the savings problem in Benhabib et al. (2015) from an unbounded problem—with unboundedness caused by the kinds of unbounded utility functions routinely used in economic modeling—into a bounded one. In a similar vein, we use a timing shift to extend the optimality results for risk sensitive preferences presented in Bäuerle and Jaśkiewicz (2018) to a larger class of models, including those with standard utility functions that are unbounded below.

Perhaps the closest result to the theoretical component of our work in the economic literature is theorem 3.2 of Rust (1994), which discusses the connection between the fixed point of a modified Bellman equation and optimality of the policy that results from choosing the maximal action at each state evaluated according to this fixed point. This result is, however, specific to one specialized class of dynamic programming problems, with discrete choices and additively separable preferences, and refers to one specific plan factorization associated with the expected value function. In contrast, our result considers all dynamic programming problems in economics of which we are aware, and can accommodate all existing plan factorizations and a range of new ones. We also consider sufficient conditions for optimality, such as contractivity and monotonicity of the modified Bellman operator, rates of convergence, and the way that these optimality conditions interact with a wide range of algorithms. Finally, our paper contains new applications of plan factorizations, as discussed above.

Some related ideas can also be found outside of economics as well. For example, the Q-learning algorithm (see, e.g., Kochenderfer (2015)) is built around a functional equation for "state-action" value, which computes lifetime value conditional on current action and subsequent optimal behavior, including both current and expected future rewards. The key functional equation is special case of what is covered here within the concept of plan factorizations (although the Q-learning algorithm itself is significantly different, in that the assumed knowledge of the controller regarding state transitions and rewards is much lower than the settings we consider.) The method of integrating out "uncontrollable states" (see, e.g., Bertsekas (2017), section 1.4) is related to the technique used in Winberry (2018). This is another special case of the decompositions treated below, corresponding to a specific time change and additively separable preferences.

2. General Formulation

This section presents an abstract dynamic programming problem and the key concepts and operators related to plan factorization. Moreover, fundamental properties regarding fixed points and iterations of the key operators are derived.

2.1. **Problem Statement.** Throughout we use $\mathbb{N}$ to denote the set of natural numbers and $\mathbb{N}_0 := \{0\} \cup \mathbb{N}$. Let $\mathbb{R}^{\mathsf{X}}$ be the set of real-valued functions defined on X. In what follows, a dynamic programming problem consists of

- a set X called the *state space*,
- a set A called the *action space*,
- a nonempty correspondence Γ from X to A called the *feasible correspondence*, along with the associated set of *state-action pairs*

$$\mathsf{F} := \{(x, a) \in \mathsf{X} \times \mathsf{A} : a \in \Gamma(x)\},$$

- a subset $\mathbb{V}$ of $\mathbb{R}^{\mathsf{X}}$ called the set of *candidate value functions*,
- a *state-action aggregator* Q mapping $\mathsf{F} \times \mathbb{V}$ to $\mathbb{R} \cup \{-\infty\}$.

The interpretation is that $Q(x, a, v)$ is the lifetime value associated with choosing action a at current state x and then continuing with a reward function v attributing value to states. In other words, $Q(x, a, v)$ is an abstract representation of the value to be maximized on the right hand side of the Bellman equation. This abstract representation accommodates a wide variety of separable and nonseparable preferences.[8]

Example 2.1. If we take the Bellman equation

$$v(w) = \max_{0 \leqslant c \leqslant w} \left\{ u(c) + \beta \mathbb{E}\, v(R'(w - c) + y') \right\}$$

from Benhabib et al. (2015), where w is wealth, c is current consumption, R' and y' are IID draws of return on assets and nonfinancial income respectively, then, assuming the household cannot borrow, $\mathsf{X} = \mathsf{A} = \mathbb{R}_+$, the feasible correspondence is $\Gamma(w) = [0, w]$, the set $\mathbb{V}$ is some subset of the continuous functions from $\mathbb{R}_+$ to $\mathbb{R} \cup \{-\infty\}$ and

$$Q(w, c, v) := u(c) + \beta \mathbb{E}\, v(R'(w - c) + y').$$

[8] Our abstract framework is motivated by the treatment of dynamic programming in Bertsekas (2013). Although the presentation here and below focuses on stationary infinite horizon problems, nonstationary and finite horizon problems can also be incorporated by including time as a state variable and allowing for nonstationary policies.

To shift to Epstein–Zin preferences, as in, say, Kaplan and Violante (2017), we can adjust Q to

$$Q(w, c, v) = \left\{ (1-\beta) u(c)^{\rho} + \beta \left[\mathbb{E}\, v(R'(w-c) + y')^{\alpha} \right]^{\rho/\alpha} \right\}^{1/\rho},$$

where α and ρ are parameters.

Returning to the general case, we can associate to our abstract dynamic programming problem the Bellman equation

$$v(x) = \max_{a \in \Gamma(x)} Q(x, a, v) \quad \text{for all } x \in \mathsf{X}. \tag{1}$$

Stating that $v \in \mathbb{V}$ solves the Bellman equation is equivalent to stating that v is a fixed point of the Bellman operator, in this case given by

$$T\, v(x) := \max_{a \in \Gamma(x)} Q(x, a, v).$$

2.2. **Policies and Assumptions.** Let Σ denote the set of *feasible policies*, which we define as all $\sigma\colon \mathsf{X} \to \mathsf{A}$ satisfying $\sigma(x) \in \Gamma(x)$ for all $x \in \mathsf{X}$ and

$$v \in \mathbb{V} \text{ and } w(x) = Q(x, \sigma(x), v) \text{ on } \mathsf{X} \implies w \in \mathbb{V}. \tag{2}$$

Given $v \in \mathbb{V}$, a feasible policy σ with the property that

$$Q(x, \sigma(x), v) = \sup_{a \in \Gamma(x)} Q(x, a, v) \quad \text{for all } x \in \mathsf{X} \tag{3}$$

is called *v-greedy*. In other words, a v-greedy policy is one we obtain by treating v as the value function and maximizing the right hand side of the Bellman equation. The following assumption allows us to work with maximal decision rules, rather than suprema. Combined with (2), it also implies that T maps $\mathbb{V}$ into itself.

Assumption 2.1. At least one v-greedy policy exists in Σ for each v in $\mathbb{V}$.

Given $\sigma \in \Sigma$, any function v_σ in $\mathbb{V}$ satisfying

$$v_\sigma(x) = Q(x, \sigma(x), v_\sigma) \quad \text{for all } x \in \mathsf{X}$$

is called a *σ-value function*. We understand $v_\sigma(x)$ as the lifetime value of following policy σ now and forever, starting from current state x.

Assumption 2.2. For each $\sigma \in \Sigma$, there is exactly one σ-value function in $\mathbb{V}$, denoted in what follows by v_σ.

Assumption 2.2 is simple to verify for regular Markov decision processes with bounded rewards and discount factors in $(0, 1)$ using, say, Banach's contraction mapping theorem or the Neumann series lemma. At the same time, it can be difficult to establish in models where rewards are unbounded or dynamic preferences are specified recursively.[9] Nevertheless, assumption 2.2 is a minimal requirement because when it fails the fundamental objective of dynamic programming—that is, maximization of lifetime value—is not well defined.

2.3. **Decompositions and Plan Factorizations.** Next we introduce plan factorizations in a relatively abstract form, motivated by the desire to accommodate all existing transformations and admit new ones, both for additively separable and for recursive preference models. As a start, let $\mathbb{H}$ be the set of all functions h in $\mathbb{R}^{\mathsf{F}}$ such that, for some $v \in \mathbb{V}$ we have $h(x, a) = Q(x, a, v)$ for all $(x, a) \in \mathsf{F}$ and let M be the operator defined at $h \in \mathbb{H}$ by

$$(Mh)(x) = \max_{a \in \Gamma(x)} h(x, a). \tag{4}$$

In the sequel, a *plan factorization* associated with the dynamic program described above is a pair of operators (W_0, W_1) such that

(a) W_0 is defined on $\mathbb{V}$ and takes values in $\mathbb{G} := W_0 \mathbb{V}$,[10]
(b) W_1 is defined on $\mathbb{G}$ and takes values in $\mathbb{H}$ and
(c) the composition $W_0 \circ W_1$ satisfies

$$(W_1 \, W_0 \, v)(x, a) = Q(x, a, v) \quad \text{for all } (x, a) \in \mathsf{F}, \; v \in \mathbb{V}. \tag{5}$$

Equation (5) states that W_0 and W_1 provide a decomposition (or factorization) of Q, so that each element W_i implements a separate component of the two stage (i.e., present and future) planning problem associated with the Bellman equation.

Evidently, given the definition of M in (4) and the factorization requirement (5), for each plan factorization (W_0, W_1), the Bellman operator T can be decomposed as

$$T = M \circ W_1 \circ W_0. \tag{6}$$

[9] See, for example, Epstein and Zin (1989), Bäuerle and Jaśkiewicz (2018), Marinacci and Montrucchio (2010) or Bloise and Vailakis (2018). The restrictions on primitives used to obtain the existence and uniqueness requirement in assumption 2.2 vary substantially across applications, which is why we have chosen to assume it directly.

[10] More explicitly, $\mathbb{G}$ is the set of $g \in \mathbb{R}^{\mathsf{F}}$ such that $g = W_0 \, v$ for some $v \in \mathbb{V}$.

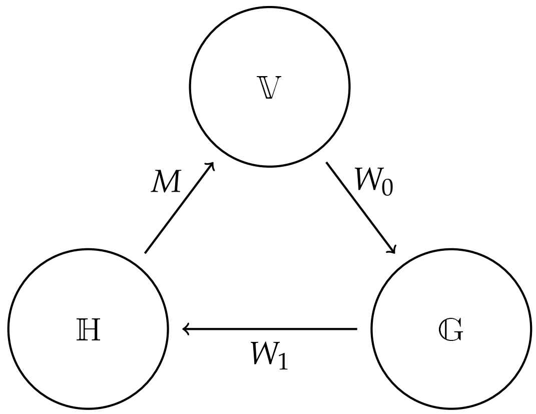

FIGURE 1. The one-shift operators

A visualization is given in figure 1, where W_0, W_1 and M are clockwise rotations and T is a cycle starting from $\mathbb{V}$.

Corresponding to this same plan factorization (W_0, W_1), we introduce a *refactored Bellman operator* S on $\mathbb{G}$ defined by

$$S = W_0 \circ M \circ W_1 \tag{7}$$

In figure 1, S is a cycle starting from $\mathbb{G}$. Corresponding to S, we have the *refactored Bellman equation* $g = W_0 \, M \, W_1 \, g$. Evidently fixed points of S are solutions to the refactored Bellman equation and vice versa. The refactored Bellman operator and refactored Bellman equation are possible implementations of an action-contingent value transformation, in the sense that they manipulate the Bellman operator and Bellman equation while attempting to preserve optimality of the policy they produce. In fact preservation of optimality requires additional assumptions, to be discussed below.[11]

2.4. **Examples.** This section helps illustrate how plan factorizations and the refactored Bellman operators they generate can represent the kinds of manipulations of

[11]In some dynamic programs, the controller has the option of stopping or exiting the decision loop in some states or actions. Such state-action pairs have the property that the continuation value does not depend on the candidate value function. We can treat them efficiently by introducing a subset F_e of F and a function $e \colon \mathsf{F}_e \to \mathbb{R}$ such that $Q(x, a, v) = e(x, a)$ for all $v \in \mathbb{V}$ and $(x, a) \in \mathsf{F}_e$. It is immediate from the definitions of the two Bellman operators that their images depend only on primitives on the set F_e. Hence we can ignore these states during iteration of any of these operators, reinserting them only when greedy policies are calculated.

the Bellman equation currently used in economic modeling. (We focus on relatively simple cases, with more complex examples deferred to sections 4–5.)

2.4.1. *Optimal Savings.* Consider a household savings problem such as found in, say, Krusell and Smith (1998), with Bellman equation

$$v(w, z, \eta) = \max_{c, w', \ell} \left\{ u(c, \ell) + \beta \, \mathbb{E}_z \, v(w', z', \eta') \right\} \tag{8}$$

subject to

$$w' + c \leqslant R(z)w + q(z, \eta)\ell, \quad c, w' \geqslant 0 \quad \text{and} \quad 0 \leqslant \ell \leqslant 1. \tag{9}$$

Here w is wealth, $R(z)$ is a gross rate of return on financial assets, c is consumption, ℓ is labor, $\{\eta_t\}$ are IID innovations, $\{z_t\}$ is a Markov state process and $\mathbb{E}_z$ conditions on current state z. The constraint set $\Gamma(w, z, \eta)$ is all c, w', ℓ satisfying (9). The state-action aggregator Q for this model is

$$Q(w, z, \eta, w', c, \ell) = u(c, \ell) + \beta \, \mathbb{E}_z \, v(w', z', \eta'). \tag{10}$$

We can use the law of iterated expectations to rewrite the Bellman equation as

$$v(w, z, \eta) = \max_{c, w', \ell} \left\{ u(c, \ell) + \beta \, \mathbb{E}_z \, \mathbb{E}_{z'} \, v(w', z', \eta') \right\}. \tag{11}$$

Next, analogous to Winberry (2018), we set

$$g(w, z) := \mathbb{E}_z \, v(w, z, \eta) \tag{12}$$

and rearrange (11) to eliminate v, producing the modified Bellman equation

$$g(w, z) = \mathbb{E}_z \max_{c, w', \ell} \left\{ u(c, \ell) + \beta \mathbb{E}_z \, g(w', z') \right\}, \tag{13}$$

with the same constraint (9). The transient shock η is integrated out by the outer expectation in (13), implying that g in (13) acts on a space one dimension lower than v in (8).

Now let us connect (13) to the theoretical framework presented above. To this end, consider the plan factorization (W_0, W_1) defined by

$$(W_0 \, v)(w, z) := \mathbb{E}_z \, v(w, z, \eta) \quad \text{and} \quad (W_1 \, g)(w', z, c, \ell) := u(c, \ell) + \beta \mathbb{E}_z g(w', z').$$

Evidently W_0 and W_1 factorize Q from (10) in the sense of condition (5), since, by construction,

$$(W_1 \, W_0 \, v)(w, z, \eta, w', c, \ell) = u(c, \ell) + \beta \, \mathbb{E}_z \mathbb{E}_{z'} \, v(w', z', \eta') = Q(w, z, \eta, w', c, \ell).$$

In addition, using the definition of M in (4), we have

$$(M\, W_1\, g)(w, z, \eta) = \max_{c, w', \ell} \left\{ u(c, \ell) + \beta \mathbb{E}_z\, g(w', z') \right\},$$

where the dependence on η is due to the constraint (9). Applying W_0 to this last expression gives (13). In other words, the modified Bellman equation (13) can be expressed as $g = W_0\, M\, W_1\, g$. Hence, g solves (13) if and only if g is a fixed point of the refactored Bellman operator $S = W_0 \circ M \circ W_1$ induced by the plan factorization (W_0, W_1).

2.4.2. *Job Search.* Consider a version of the McCall (1970) job search model with IID wage offers $\{w_t\}$ and unemployment compensation $\{\eta_t\}$. Ignoring complications such as job separation, the Bellman equation for the agent is

$$v(w, \eta) = \max \left\{ \frac{u(w)}{1-\beta},\, u(\eta) + \beta \mathbb{E}\, v(w', \eta') \right\}. \tag{14}$$

The state-action aggregator Q for this model can be expressed as

$$Q(w, \eta, a) = a \frac{u(w)}{1-\beta} + (1-a) \left[u(\eta) + \beta \mathbb{E}\, v(w', \eta') \right]$$

where $a \in \Gamma(w, \eta) := \{0, 1\}$ is a binary choice variable, with $a = 1$ indicating the decision to stop.

It is common to transform the Bellman equation for this problem using the continuation value $g(\eta) := u(\eta) + \beta \mathbb{E}\, v(w', \eta')$. Using g to eliminate v from (14) gives

$$g(\eta) = u(\eta) + \beta \mathbb{E} \max \left\{ \frac{u(w')}{1-\beta},\, g(\eta') \right\},$$

a functional equation in one dimension. We can in fact do better by using the expected value function, analogous to the technique used by Rust (1987) in the context of regenerative optimal stopping. In particular, we can take $g := \mathbb{E}\, v(w', \eta')$ and eliminate v from the Bellman equation to obtain the scalar equation

$$g = \mathbb{E} \max \left\{ \frac{u(w')}{1-\beta},\, u(\eta') + \beta g \right\}. \tag{15}$$

To frame this transformation in terms of the refactored Bellman operator, consider the plan factorization (W_0, W_1) defined by

$$W_0\, v = \mathbb{E}\, v(w', \eta') \quad \text{and} \quad (W_1\, g)(\eta, w, a) = a \frac{u(w)}{1-\beta} + (1-a) \left[u(\eta) + \beta g \right]. \tag{16}$$

Notice that the functional equation (15) can be rewritten as

$$g = \mathbb{E}\left\{\max_{a\in\{0,1\}}\left\{a\,\frac{u(w')}{1-\beta} + (1-a)[u(\eta') + \beta g]\right\}\right\}, \tag{17}$$

which, with the definitions of W_0 and W_1 just given, is equivalent to $g = W_1\, M\, W_0\, g$. Thus, a scalar g solves (15) if and only if it is a fixed point of the refactored Bellman operator $S = W_0 \circ M \circ W_1$ generated by the factorization in (16).

2.5. **Refactored Policy Values.** Returning to the general setting, we also wish to consider the value of policies under the transformation associated with a given plan factorization (W_0, W_1). To this end, for each $\sigma \in \Sigma$ and each $h \in \mathbb{H}$, we define the operator $M_\sigma \colon \mathbb{H} \to \mathbb{V}$ by

$$M_\sigma h(x) := h(x, \sigma(x)). \tag{18}$$

That M_σ does in fact map $\mathbb{H}$ into $\mathbb{V}$ follows from the definition of these spaces and (2).[12] Comparing with M defined in (4), we have, for each $h \in \mathbb{H}$

$$M_\sigma h \leqslant Mh \quad \text{and} \quad M_\sigma h = Mh \text{ for at least one } \sigma \in \Sigma. \tag{19}$$

The inequality in (19) is obvious. To see that the existence claim holds, observe that, by the definition of $\mathbb{H}$, there exists a $v \in \mathbb{V}$ such that $h = W_1 W_0 v$. For this v, by assumption 2.1, there exists a σ in Σ such that $M_\sigma\, W_1\, W_0\, v = M\, W_1\, W_0\, v$. Since $h = W_1\, W_0\, v$, this verifies the second claim in (19).

Given $\sigma \in \Sigma$, the operator T_σ from $\mathbb{V}$ to itself defined by $T_\sigma v(x) = Q(x, \sigma(x), v)$ for all $x \in \mathsf{X}$ will be called the *σ-value operator*. By construction, it has the property that v_σ is the σ-value function corresponding to σ if and only if it is a fixed point of T_σ. With the notation introduced above, we can express it as

$$T_\sigma = M_\sigma \circ W_1 \circ W_0$$

Analogous to the definition of the refactored Bellman operator in (7), we introduce the *refactored σ-value operator*

$$S_\sigma := W_0 \circ M_\sigma \circ W_1$$

corresponding to a given plan factorization (W_0, W_1). A fixed point g_σ of S_σ is called a *refactored σ-value function*. The value $g_\sigma(x, a)$ can be interpreted as the value of following policy σ in all subsequent periods, conditional on current action a.

[12] If $h \in \mathbb{H}$, then by definition there exists a $v \in \mathbb{V}$ such that $h(x, a) = Q(x, a, v)$ for all $(x, a) \in \mathsf{F}$. Now $M_\sigma h \in \mathbb{V}$ follows directly from (2).

2.6. **Fixed Points and Iteration.** We begin with some preliminary results not directly connected to optimality. Our first result states that, to iterate with T, one can alternatively iterate with S, since iterates of S can be converted directly into iterates of T. Moreover, the converse is also true, and similar statements hold for the policy operators:

Proposition 2.1. *For every $n \in \mathbb{N}$ we have*

$$S^n = W_0 \circ T^{n-1} \circ M \circ W_1 \quad \textit{and} \quad T^n = M \circ W_1 \circ S^{n-1} \circ W_0.$$

Moreover, for every $n \in \mathbb{N}$ and every $\sigma \in \Sigma$, we have

$$S^n_\sigma = W_0 \circ T^{n-1}_\sigma \circ M_\sigma \circ W_1 \quad \textit{and} \quad T^n_\sigma = M_\sigma \circ W_1 \circ S^{n-1}_\sigma \circ W_0.$$

Proposition 2.1 follows from the definitions of T and S along with a simple induction argument. While the proof is straightforward, the result is worth bearing in mind because it provides a metric-free statement of the fact that the sequence of iterates of any refactored Bellman operator converges at the same rate as that of the standard Bellman operator.

The next result shows a fundamental connection between the two forms of the Bellman operator in terms of their fixed points.

Proposition 2.2. *The Bellman operator T admits a unique fixed point $\bar{v}$ in $\mathbb{V}$ if and only if the refactored Bellman operator S admits a unique fixed point $\bar{g}$ in $\mathbb{G}$. Whenever these fixed points exist, they are related by $\bar{v} = M\, W_1\, \bar{g}$ and $\bar{g} = W_0\, \bar{v}$.*

Thus, if a unique fixed point of the Bellman operator is desired but T is difficult to work with, a viable option is to show that S has a unique fixed point, compute it, and then recover the unique fixed point of T via $\bar{v} = M\, W_1\, \bar{g}$.

The fact that T has a unique fixed point in $\mathbb{V}$ does not by itself allow us to draw any conclusions about optimality. A discussion of optimality is provided in section 3.

A result analogous to proposition 2.2 holds for the policy operators:

Proposition 2.3. *Given $\sigma \in \Sigma$, the refactored σ-value operator S_σ has a unique fixed point g_σ in $\mathbb{G}$ if and only if T_σ has a unique fixed point v_σ in $\mathbb{V}$. Whenever these fixed points exist, they are related by $v_\sigma = M_\sigma\, W_1\, g_\sigma$ and $g_\sigma = W_0\, v_\sigma$.*

Proposition 2.3 implies that, under assumption 2.2, there exists exactly one refactored σ-value function g_σ in $\mathbb{G}$ for every $\sigma \in \Sigma$. Proposition 2.3 is also useful for the converse implication. In particular, to establish assumption 2.2, which is often nontrivial when preferences are not additively separable, we can equivalently show that S_σ has a unique fixed point in $\mathbb{G}$. An application of this idea that illustrates its value is given in section 4.

3. Optimality

Next we turn to optimality, with a particular focus on the properties that must be placed on a given plan factorization in order for the corresponding (refactored) Bellman equation to lead us to optimal actions. Our first step, however, is to define optimality and recall some standard results.

3.1. **Fixed Points and Optimal Policies.** The *value function* associated with our dynamic program is defined at $x \in \mathsf{X}$ by

$$v^*(x) = \sup_{\sigma \in \Sigma} v_\sigma(x). \tag{20}$$

A feasible policy σ^* is called *optimal* if $v_{\sigma^*} = v^*$ on X. This corresponds to the usual definition when rewards are additively separable. *Bellman's principle of optimality* states that a policy is optimal if and only if it is v^*-greedy.

The next theorem is a trivial extension of foundational optimality results for dynamic decision problems. Its proof is included in the appendix for the reader's convenience. The assumptions of section 2.2 are taken to be in force.

Theorem 3.1. *The next two statements are equivalent:*

- *(a) The value function v^* lies in $\mathbb{V}$ and satisfies the Bellman equation.*
- *(b) Bellman's principle of optimality holds and the set of optimal policies is nonempty.*

It is natural to ask whether a result analogous to theorem 3.1 holds for the refactored Bellman operator S. Indeed, the entire motivation behind the transformations in question is that the Bellman equation can be refactored into a more convenient form without affecting optimality. By the last statement we mean that the optimal policy remains accessible through a version of Bellman's principle of optimality (i.e., computing the maximizing action associated with the right hand side of the refactored Bellman equation at each point in the state space produces the optimal policy).

We now show that, while some care in treating this problem is required, a result analogous to theorem 3.1 can be established if a form of monotonicity holds for the transformation being used in the plan factorization. Before we get to this result, however, we need to address the following complication: there are two distinct functions that can take the part of v^* theorem 3.1. One is the rotated value function

$$\hat{g} := W_0 \, v^*. \tag{21}$$

In general, $\hat{g}$ is a function of both state x and action a.

Example 3.1. In the optimal savings problem of section 2.4.1, we considered the integrated value function $g(w,z) := \mathbb{E}_z \, v(w,z,\eta)$, which was expressed in terms of a plan factorization as $g(w,z) = (W_0 \, v)(w,z)$. If $v = v^*$, the value function associated with this optimal savings problem, then $\hat{g}(w,z) = (W_0 \, v^*)(w,z)$.

The second function we need to consider as a possible fixed point of S is

$$g^*(x,a) := \sup_{\sigma \in \Sigma} g_\sigma(x,a). \tag{22}$$

We call g^* the *refactored value function*. The definition of g^* directly parallels the definition of the value function in (20), with $g^*(x,a)$ representing the maximal value that can be obtained from state x conditional on choosing a in the current period. In this sense it is more consequential than $\hat{g}$, which is important only because it is obtained by rotating v^*. (Since assumption 2.2 is in force, the set of functions $\{g_\sigma\}$ in the definition of g^* is well defined (see proposition 2.3), and hence so is g^* as an extended real-valued function, although it might or might not live in $\mathbb{G}$.)

As shown below, the functions $\hat{g}$ and g^* are not in general equal, although they become so when a certain form of monotonicity is imposed. Moreover, under this same monotonicity condition, if one and hence both of these functions are fixed points of S, we obtain valuable optimality results.

In stating these result, we recall that a map A from one partially ordered set $(E, \leqslant)$ to another such set $(F, \leqslant)$ is called *isotone* if $x \leqslant y$ implies $Ax \leqslant Ay$. Below, isotonicity is with respect to the pointwise partial order on $\mathbb{V}$, $\mathbb{G}$ and $\mathbb{H}$. Also, given $g \in \mathbb{G}$, a policy $\sigma \in \Sigma$ is called g-*greedy* if $M_\sigma \, W_1 \, g = M \, W_1 \, g$. As a result of assumption 2.1, at least one such policy exists for every $g \in \mathbb{G}$.[13]

[13]By definition, $g \in \mathbb{G}$ implies the existence of a $v \in \mathbb{V}$ such that $g = W_0 \, v$, and to this v there corresponds a v-greedy policy σ by assumption 2.1. At this σ we have $Q(x, \sigma(x), v) = \max_{a \in \Gamma(x)} Q(x,a,v)$ for every $x \in \mathsf{X}$, or, equivalently, $M_\sigma \, W_1 \, W_0 \, v = M \, W_1 \, W_0 \, v$ pointwise on $\mathbb{V}$. Since $g = W_0 \, v$, this policy is g-greedy.

Theorem 3.2. *Let (W_0, W_1) be a plan factorization, let $\hat{g}$ be as defined in (21) and let g^* be as defined in (22). If W_0 and W_1 are both isotone, then the following statements are equivalent:*

(a) g^* lies in $\mathbb{G}$ and satisfies the refactored Bellman equation.
(b) v^* lies in $\mathbb{V}$ and satisfies the Bellman equation.

If either one of these conditions holds, then $g^* = \hat{g}$, the set of optimal policies is nonempty and, for $\sigma \in \Sigma$, we have

$$\sigma \text{ is optimal} \iff \sigma \text{ is } g^*\text{-greedy} \iff \sigma \text{ is } v^*\text{-greedy.}$$

Example 3.2. Observe that the operators W_0 and W_1 that form the plan factorization for the optimal savings problem considered in section 2.4.1 are both isotone, so if we can show that the refactored value function g^* satisfies (13), then by theorem 3.2 at least one optimal policy exists, $g^*(w, z) = \hat{g}(w, z) = \mathbb{E}_z v^*(w, z, \eta)$, and a policy $\sigma \in \Sigma$ is optimal if and only if

$$\sigma(w, z) \in \operatorname*{argmax}_{c, w', \ell} \left\{ u(c, \ell) + \beta \mathbb{E}_z g^*(w', z') \right\}$$

subject to the constraint (9) for all $(w, z) \in \mathbb{R}^2_+$.

More sophisticated applications of theorem 3.2 are deferred to sections 4–5. Also, theorem 3.2 does not address the issue of how to establish that g^* satisfies the refactored Bellman equation. This problem is considered in section 3.3.

3.2. **The Role of Monotonicity.** Isotonicity of W_0 and W_1 cannot be dropped from theorem 3.2 without changing its conclusions. In section 6.5 of the appendix we exhibit dynamic programs and plan factorizations that illustrate the following possibilities:

(a) $\hat{g} \neq g^*$
(b) $Tv^* = v^*$ and yet $Sg^* \neq g^*$
(c) $Sg^* = g^*$ and yet $Tv^* \neq v^*$

Our examples are relatively stylized and in practice the isotonicity assumptions are typically satisfied. At the same time, transformations where isotonicity fails are not difficult to envisage. For example, consider a risk-sensitive control problem with Bellman operator

$$Tv(x) = \max_{a \in \Gamma(x)} \left\{ r(x,a) - \frac{\beta}{\gamma} \ln \int \exp(-\gamma v(F(x,a,z)))\mu(\mathrm{d}z) \right\}$$

as in, say, Hansen and Sargent (2008) or Bäuerle and Jaśkiewicz (2018). Here r is a current reward function and $F(x,a,z)$ is a transition rule conditional on state x, action a and shock z (with distribution μ), while β is a discount factor and $\gamma > 0$ controls the degree of risk sensitivity. If, in this context, we use an "integrated value function" of the form

$$g(x,a) := (W_0\, g)(x,a) := \int \exp(-\gamma v(F(x,a,z)))\mu(\mathrm{d}z)$$

then W_0 is antitone rather than isotone. In contrast, if we use

$$g(x,a) := (W_0\, g)(x,a) := -\frac{1}{\gamma} \ln \int \exp(-\gamma v(F(x,a,z)))\mu(\mathrm{d}z)$$

then isotonicity holds. In fact this transformation can produce valuable results, as shown in section 5.

3.3. **Sufficient Conditions.** Theorem 3.2 tells us that if the stated monotonicity condition holds and $Sg^* = g^*$, then we can be assured of the existence of optimal policies and have a means to characterize them. What we lack is a set of sufficient conditions under which the refactored Bellman operator has a unique fixed point that is equal to g^*.

To study this issue, we recall that a self-mapping A on a topological space U is called *asymptotically stable* on U if A has a unique fixed point $\bar{u}$ in U and $A^n u \to \bar{u}$ as $n \to \infty$ for all $u \in U$. In the theorem below, (W_0, W_1) is a given plan factorization and g^* is the refactored value function.

Theorem 3.3. *Suppose there exists a Hausdorff topology τ on $\mathbb{G}$ under which the pointwise partial order is closed and $\{S, S_\sigma\}_{\sigma \in \Sigma}$ are all asymptotically stable on $\mathbb{G}$.*[14] *If, in addition, W_0 and W_1 are both isotone, then*

(a) g^* *is the unique solution to the refactored Bellman equation in* $\mathbb{G}$,
(b) $\lim_{k \to \infty} S^k g = g^*$ *under* τ *for all* $g \in \mathbb{G}$,

[14] The pointwise partial order $\leqslant$ is closed with respect to τ if its graph closed in the product topology on $\mathbb{G} \times \mathbb{G}$. The key implication for us is that if $f_n \to f$ and $g_n \to g$ under τ and $f_n \leqslant g_n$ for all n, then $f \leqslant g$.

(c) at least one optimal policy exists and

(d) a feasible policy is optimal if and only if it is g^*-greedy.

Thus, under the conditions of theorem 3.3, the refactored Bellman equation does indeed have "equal rights." We can (in the limit) compute the refactored value function g^* by iterating with this operator, and the policy produced by taking the maximizer associated with the max operation inside the refactored Bellman equation evaluated at g^*—what we call the g^*-greedy policy—is optimal.

In theorem 3.3 we have eshewed a contractivity assumption on S or S_σ, even though contraction map arguments are often used in this setting, since contractivity is problematic in some applications of interest (consider the LQ regulator problem or see the discussion of recursive preferences in Marinacci and Montrucchio (2010) or Bloise and Vailakis (2018)). At the same time, contraction methods can in fact be applied to many problems, and availability of the refactored Bellman operator opens up more. For this reason we add the next proposition, in the statement of which the map $\|\cdot\|_\kappa$ is defined at $f \in \mathbb{R}^{\mathsf{F}}$ by $\|f\|_\kappa = \sup |f/\kappa|$. Here the supremum is over all $(x, a) \in \mathsf{F}$ and κ is a fixed "weighting function;" that is, an element of $\mathbb{R}^{\mathsf{F}}$ satisfying $\kappa \geqslant 1$ on F. This mapping defines a norm on $b_\kappa \mathsf{F}$, the set of $f \in \mathbb{R}^{\mathsf{F}}$ such that $\|f\|_\kappa < \infty$. The metric induced by this norm is complete and the pointwise order is closed under the norm topology.

Proposition 3.4. *Let W_0 and W_1 be isotone. If, for some weighting function κ on F, the set $\mathbb{G}$ is a closed subset of $b_\kappa \mathsf{F}$ and there exists a positive constant β such that $\beta < 1$ and*

$$\|S_\sigma g - S_\sigma g'\|_\kappa \leqslant \beta \, \|g - g'\|_\kappa \tag{23}$$

for all $g, g' \in \mathbb{G}$ and $\sigma \in \Sigma$, then S and $\{S_\sigma\}_{\sigma \in \Sigma}$ are asymptotically stable on $\mathbb{G}$ and the conclusions of theorem 3.3 hold. The sequence $S^k g$ converges to g^* at rate $O(\beta^k)$ under $\|\cdot\|_\kappa$.

Notice that the contractivity requirement is imposed on S_σ rather than S, making it easier to verify (since S_σ does not involve a maximization step). An application of proposition 3.4 is given in section 4.

3.4. **Refactored Policy Iteration.** Next we give conditions under which modified policy iteration (sometimes called optimistic policy iteration) is successful in the setting of a given plan factorization (W_0, W_1). The standard algorithm starts with an initial candidate $v_0 \in \mathbb{V}$ and generates sequences $\{\sigma_k\}$ and $\{\Sigma_k\}$ in Σ and $\{v_k\}$ in $\mathbb{V}$ by taking

$$\sigma_k \in \Sigma_k \quad \text{and} \quad v_{k+1} = T_{\sigma_k}^{m_k} v_k \quad \text{for all } k \in \mathbb{N}_0, \tag{24}$$

where Σ_k is the set of v_k-greedy policies, and $\{m_k\}$ is a sequence of positive integers. The first step of equation (24) is called *policy improvement*, while the second step is called *partial policy evaluation*. If $m_k = 1$ for all k then the algorithm reduces to value function iteration.

To extend this idea to the refactored case, we take an initial candidate $g_0 \in \mathbb{G}$ and generate sequences $\{\sigma_k\}$ and $\{\Sigma_k\}$ in Σ and $\{g_k\}$ in $\mathbb{G}$ via

$$\sigma_k \in \Sigma_k \quad \text{and} \quad g_{k+1} = S_{\sigma_k}^{m_k} g_k \quad \text{for all } k \in \mathbb{N}_0, \tag{25}$$

where Σ_k is the set of g_k-greedy policies, and $\{m_k\}$ is a sequence of positive integers.

The next result shows that (24) and (25) indeed generate the same sequences of sets of greedy policies.

Proposition 3.5. *If $v_0 \in \mathbb{V}$ and $g_0 = W_0 v_0$, then the sequences $\{\sigma_k\}$ and $\{\Sigma_k\}$ satisfy (24) if and only if they satisfy (25). Furthermore, $g_k = W_0 v_k$ for all k.*

The next theorem pertains to a sequence generated by (25). In the statement of the theorem, asymptotic stability is with respect to a topology τ on $\mathbb{G}$ under which the pointwise partial order is closed.

Theorem 3.6. *Let W_0 and W_1 be isotone and let S and $\{S_\sigma\}_{\sigma \in \Sigma}$ be asymptotically stable on $\mathbb{G}$. If $g_0 \leqslant S g_0$, then $g_k \leqslant g_{k+1}$ for all k and $g_k \to g^*$ as $k \to \infty$.*

4. Application Set I: Unbounded Rewards

One of the more problematic aspects of dynamic programmingn is dealing with rewards that are unbounded—and particularly those applications were rewards are unbounded below. At the same time, such specifications abound in economics. In this section, we address these problems using a new approach based on plan factorization. We show that, in certain applications, while the Bellman operator acts on a space of functions that are unbounded below, the refactored Bellman operator sends functions that are bounded below into the same class of functions. We use our results to

(a) show the validity of (refactored) value function iteration and modified policy iteration for the optimal savings model studied in Benhabib et al. (2015), and

(b) extend the optimality results for risk sensitive preference models obtained in Bäuerle and Jaśkiewicz (2018) to the case where rewards are unbounded below.

4.1. **Theory.** Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and let $L_1 = L_1(\Omega, \mathscr{F}, \mathbb{P})$ be all random variables X on Ω with $\mathbb{E}\,|X| < \infty$. Let $\mathcal{R}$ be a certainty equivalent operator, by which we mean a map from L_1 to $\mathbb{R}$ satisfying $\mathcal{R}\alpha = \alpha$ for any constant α and $\mathcal{R}X \leqslant \mathcal{R}X'$ whenever $X, X' \in L_1$ with $X \leqslant X'$. Suppose in addition that

(P1) $\mathcal{R}X \leqslant \mathbb{E}\,X$ for all $X \in L_1$, and

(P2) $\mathcal{R}(X + X') \leqslant \mathcal{R}X + \mathcal{R}X'$ for all $X, X' \in L_1$.

Here P1 indicates risk aversion while P2 is called sub-additivity. These properties are satisfied by some but not all certainty equivalent operators.[15]

Now consider a dynamic programming problem with state-action aggregator

$$Q(x, a, v) = r(x, a) + \beta \mathcal{R} v[f(a, \xi)]. \tag{26}$$

In particular, $\mathsf{A} = \mathsf{X} = \mathbb{R}_+$, $\Gamma(x) = [0, x]$, the reward function r is defined by

$$r : D \to \mathbb{R} \cup \{-\infty\} \quad \text{where} \quad D := \{(x, a) \in \mathsf{X} \times \mathsf{A} : a \leqslant x\},$$

the function f maps $\mathbb{R}_+^2$ to $\mathbb{R}_+$, ξ is a nonnegative random variable with distribution μ and $\beta \in (0, 1)$. The Bellman operator corresponding to this problem is

$$Tv(x) = \max_{0 \leqslant a \leqslant x} \left\{ r(x, a) + \beta \mathcal{R} v[f(a, \xi)] \right\}. \tag{27}$$

Assumption 4.1. The primitives $(f, \xi, r, \mathcal{R})$ satisfy the following conditions:

(a) r is increasing in its first argument and decreasing in its second.

(b) Either r is continuous on $D_0 := \{(x, a) \in D : x > a\}$ and $\lim_{a \to x} r(x, a) = -\infty$, or r is continuous and $r(x, a) > -\infty$ on D / D_0.

(c) There exists a constant $d > 0$ and an increasing continuous function $\kappa : \mathbb{R}_+ \to [1, \infty)$ such that $r(x, 0) \leqslant d\kappa(x)$ for all $x \in \mathbb{R}_+$.

(d) The function $a \mapsto f(a, z)$ is continuous and increasing for all $z \in \mathbb{R}_+$. If $\lim_{a \to x} r(x, a) = -\infty$, then $f(0, \xi) > 0$ μ-almost surely.

[15]See, for example, Marinacci and Montrucchio (2010), Bloise and Vailakis (2018) or Bäuerle and Jaśkiewicz (2018).

(e) There exists a constant $\alpha \in (0, 1/\beta)$ such that

$$\int_{\mathbb{R}_+} \kappa[f(a,z)]\mu(\mathrm{d}z) \leqslant \alpha\kappa(a) \quad \text{for all } \; a \in \mathbb{R}_+.$$

(f) For a Borel measurable map $h\colon \mathbb{R}_+^2 \to \mathbb{R} \cup \{-\infty\}$, the function $a \to \mathcal{R}h(a,\xi)$ is Borel measurable whenever the expectation is well defined.

(g) $\mathcal{R}[r(f(0,\xi),0) + \alpha_0] \geqslant \mathcal{R}r(f(0,\xi),0) + \alpha_0 > -\infty$ for all constant α_0.

(h) If $h\colon \mathbb{R}_+^2 \to \mathbb{R} \cup \{-\infty\}$ and $a \mapsto h(a,\xi)$ is continuous μ-almost surely with $r(f(0,z),0) - \alpha_1 \leqslant h(a,z) \leqslant \alpha_2\kappa[f(a,z)]$ for some constant $\alpha_1, \alpha_2 \in \mathbb{R}_+$, then $a \mapsto \mathcal{R}h(a,\xi)$ is continuous.

Now consider the plan factorization (W_0, W_1) defined by[16]

$$(W_0\, v)(a) = \mathcal{R}v(f(a,\xi)) \quad \text{and} \quad (W_1\, g)(x,a) = r(x,a) + \beta g(a).$$

Let $\mathbb{V}$ be the set of Borel measurable functions $v : \mathsf{X} \to \mathbb{R}$ such that $W_0 v \in \mathbb{G}$, where $\mathbb{G}$ is defined as the set of Borel measurable functions $g : \mathbb{R}_+ \to \mathbb{R}$ such that

- $\|g\|_\kappa := \sup_{a \in \mathbb{R}_+}[g(a)/\kappa(a)]$ is finite and
- g is increasing, continuous and bounded below.

Lemma 4.1. *If assumption 4.1 holds,* $h_g(x) := \max_{0 \leqslant a \leqslant x}\{r(x,a) + \beta g(a)\}$ *and* $M_g(x) := \{a \in [0,x] : h_g(x) = r(x,a) + \beta g(a)\}$ *for all* $g \in \mathbb{G}$*, then*

(a) h_g *is well defined and increasing.*

(b) h_g *is continuous on* $(0,\infty)$.

(c) h_g *is continuous everywhere if* $r(x,a) > -\infty$ *on* D/D_0.

(d) M_g *is nonempty, compact-valued, and upper hemicontinuous.*

The refactored Bellman operator $S = W_0 \circ M \circ W_1$ is then

$$Sg(a) := \mathcal{R} \max_{0 \leqslant a' \leqslant f(a,\xi)} \left\{ r\left(f(a,\xi), a'\right) + \beta g(a') \right\}. \tag{28}$$

The proof of the next theorem is based on proposition 3.4.

Theorem 4.2. *If assumption 4.1 holds, then*

(a) g^* *is the unique fixed point of the refactored Bellman operator* S *in* $\mathbb{G}$*,*

[16]Clearly W_0 and W_1 factorize Q in the sense that, for Q defined in (26), we have $Q = W_1 \circ W_0$.

(b) $\lim_{k\to\infty} \|S^k g - g^*\|_\kappa = 0$ *for all* $g \in \mathbb{G}$*,*

(c) $S^k g$ *converges to* g^* *at rate* $O((\alpha\beta)^k)$ *under* $\|\cdot\|_\kappa$*,*

(d) at least one optimal policy exists and

(e) a feasible policy is optimal if and only if it is g^**-greedy.*

4.2. **Application to Optimal Savings.** Benhabib et al. (2015) studies an optimal savings problem with capital income risk, where w_t is wealth, s_t is saving, R_t is the rate of return on wealth, I_t is the labor income, and u is the utility function. As assumed by Benhabib et al. (2015), the $\{R_t\}$ and $\{I_t\}$ processes are nonnegative and IID. The Bellman equation is

$$v(w) = \max_{0 \leqslant s \leqslant w} \left\{ u(w-s) + \beta \mathbb{E}\, v(R's + I') \right\}.$$

Following Benhabib et al. (2015), we consider $u(c) = c^{1-\gamma}/(1-\gamma)$ with $\gamma > 1$.

As a dynamic program, the only difficult aspect of the problem is that utility is unbounded below, implying that the usual contraction mapping structure based on the supremum norm cannot be applied. While there are ways to circumvent this difficulty using ad hoc or specialized methods, it would be helpful if we could recast it as a bounded problem to which standard machinery can be applied. As we now show, this is possible using a plan factorization.

In particular, we consider the refactored Bellman equation

$$g(s) = \mathbb{E} \max_{0 \leqslant s' \in R's + I'} \left\{ u(R's + I' - s') + \beta g(s') \right\}. \tag{29}$$

This is a version of (28) if we make the identifications $x := w$, $a := s$ and $\xi := (R', I')$, along with $r(x,a) := u(w-s)$, $f(a,z) := R's + I'$ and $\mathcal{R} := \mathbb{E}$. Conditions (a), (b), (d), (f) of assumption 4.1 hold are easy to verify. Since u is bounded above, conditions (c), (e), (h) of assumption 4.1 hold trivially by letting $\kappa \equiv 1$. In particular, condition (h) follows from the dominated convergence theorem. Moreover, since $\mathcal{R} r(f(0,\xi),0) = \mathbb{E}\, u(I')$, assumption 4.1-(g) holds as long as $\mathbb{E}\, u(I_t) > -\infty$. Hence, by theorem 4.2, g^* is the unique fixed point of S in $\mathbb{G} := bic\mathbb{R}_+$, the space of bounded increasing continuous functions on $\mathbb{R}_+$, and $S^k g$ converges to g^* under the standard supremum norm topology at rate $O(\beta^k)$ for all $g \in \mathbb{G}$.

Next we turn to a more complex setting where optimality results are harder.

4.3. **Application to Risk Sensitive Control.** Bäuerle and Jaśkiewicz (2018) study an optimal growth model in the presence of risk sensitive preference (which are in turn related to robust control problems, as discussed in Hansen and Sargent (2008)). To represent their model in our framework, let x denote output, let a be investment, and let $r(x,a) := u(x-a)$ where u is a utility function. For all integrable $h\colon \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}$, let

$$\mathcal{R}h(a,\xi) := \phi^{-1}\left\{\mathbb{E}\,\phi[h(a,\xi)]\right\}, \quad \text{where} \quad \phi(x) = \mathrm{e}^{-\gamma x}. \tag{30}$$

Here the expectation is taken with respect to ξ. The Bellman equation is

$$v(x) = \max_{a\in[0,x]}\left\{u(x-a) - \frac{\beta}{\gamma}\ln\int_{\mathbb{R}_+}\exp\left(-\gamma\, v[f(a,z)]\right)\mu(\mathrm{d}z)\right\}.$$

Our aim is to extend the optimality results in Bäuerle and Jaśkiewicz (2018) by allowing for reward functions that are unbounded below (as well as above). We carry this out using the plan factorization discussed in section 4.1.

The refactored Bellman equation corresponding to (28) is

$$g(a) = -\frac{1}{\gamma}\ln\int_{\mathbb{R}_+}\exp\left\{-\gamma\max_{a'\in[0,f(a,z)]}\left[u\left(f(a,z)-a'\right)+\beta g(a')\right]\right\}\mu(\mathrm{d}z). \tag{31}$$

By property (P4) and proposition 1 of Bäuerle and Jaśkiewicz (2018), $\mathcal{R}$ defined in (30) is a sub-additive certainty equivalent operator. Consider, say, $u(c) = \ln c$ and

$$x_{t+1} = f(a_t, \xi_t) = \eta a_t + \xi_t, \quad \{\xi_t\} \overset{\text{IID}}{\sim} LN(0,\sigma^2),$$

where $\eta > 0$. This case is not covered by Bäuerle and Jaśkiewicz (2018) because utility is unbounded below. However, theorem 4.2 can be applied. Conditions (a), (b), (d) and (f) of assumption 4.1 obviously hold. Condition (h) follows immediately from the dominated convergence theorem. To verify condition (g), by the definition of $\mathcal{R}$, it suffices to show that

$$L := \int_{\mathbb{R}_+}\exp\{-\gamma u[f(0,z)]\}\mu(\mathrm{d}z) < \infty.$$

Since $\exp\{-\gamma u[f(0,z)]\} = \exp(-\gamma\ln z) = z^{-\gamma}$ and $\{\xi_t\}$ is lognormal, we have

$$L = \int_{\mathbb{R}_+} z^{-\gamma}\mu(\mathrm{d}z) = \exp\left(\gamma^2\sigma^2/2\right) < \infty$$

and assumption 4.1-(g) is verified. Moreover, if $\bar{z} := \int z\mu(\mathrm{d}z)$, then:

- If $\eta \leqslant 1$, then conditions (c) and (e) of assumption 4.1 hold for all $\beta \in (0,1)$ by letting $\alpha := (\beta+1)/(2\beta)$ and $\kappa(a) := a + \bar{z}/(\alpha-1)$.

- If $\eta > 1$, then conditions (c) and (e) of assumption 4.1 hold for all $\beta \in (0, 1/\eta)$ by letting $\alpha := \eta$ and $\kappa(a) := a + \bar{z}/(\alpha - 1)$.

Hence, assumption 4.1 and thus all the statements of theorem 4.2 hold.

5. Application Set II: Dimensionality Reduction

In this section we show how plan factorizations can be used to simplify a range of recent applications by reducing the dimension of the state or state-action space.

Throughout this section, we denote $\mathbb{E}_{b|a}$ as the expectation with respect to b conditional on a.

5.1. Pairwise-Difference Estimations.

Hong and Shum (2010) develop a pairwise-difference estimation method for dynamic models with discrete or continuous actions. To simplify notation and focus on fixed point solutions, we omit the parameters and agent indices of the original model. The state variable is (x_t, s_t), where x_t is observed by the econometrician and s_t is not. The agent's choice variable is q_t. The reward function is $u(x_t, s_t, q_t)$. Moreover, $\{s_t\}$ is IID, while $\{x_t\}$ evolves according to

$$x_{t+1} = x_t + q_t.$$

The Bellman operator of the problem is

$$Tv(x, s) = \max_q \left\{ u(x, s, q) + \beta \mathbb{E}_{x', s' | x, q} v(x', s') \right\}.$$

Consider the plan factorization (W_0, W_1), where

$$W_0 v(x) := \mathbb{E}_s v(x, s) \quad \text{and} \quad W_1 g(x, s, q) := u(x, s, q) + \beta \mathbb{E}_{x' | x, q} g(x').$$

The refactored Bellman operator $S = W_0 M W_1$ is then

$$Sg(x) = \mathbb{E}_s \max_q \left\{ u(x, s, q) + \beta \mathbb{E}_{x' | x, q} g(x') \right\}. \tag{32}$$

Computation via (32) only requires keeping track of the grid points of x. Once g^* is obtained, computing the optimal policy $q^*(x, s)$ is then a simple integration-plus-optimization step.

5.2. **Optimal Savings with Stochastic Discounting.** Consider the optimal savings problem studied by Higashi et al. (2009). Given constant interest rate r and current savings s_t, the agent decides current consumption c_t and savings s_{t+1} carried to the next period. The agent's discount factor α_t is stochastic and satisfies $\{\alpha_t\} \overset{\text{IID}}{\sim} \mu$. Let u be the utility function. The Bellman operator is

$$Tv(s,\alpha) = \max_{c,s'} \left\{ (1-\alpha)u(c) + \alpha \int v(s',\alpha')\mu(\mathrm{d}\alpha') \right\} \tag{33}$$

subject to

$$c + s' = (1+r)s \quad \text{and} \quad c, s' \geqslant 0. \tag{34}$$

We now consider the plan factorization (W_0, W_1), where

$$W_0 v(s) := \int v(s,\alpha)\mu(\mathrm{d}\alpha) \quad \text{and} \quad W_1 g(s,\alpha,c,s') := (1-\alpha)u(c) + \alpha g(s').$$

Then the refactored Bellman operator $S = W_0 M W_1$ is

$$Sg(s) = \int \max_{c,s'} \left\{ (1-\alpha)u(c) + \alpha g(s') \right\} \mu(\mathrm{d}\alpha) \tag{35}$$

subject to (34). Although computation via (33) is two-dimensional, computation via (35) is only unidimensional.

5.3. **Targeted Government Transfers.** Oh and Reis (2012) studies the positive effect of government transfers in a general equilibrium framework. In the household sector, k_t denotes capital, c_t is consumption, $n_t \in \{0,1\}$ is the choice to work or not, r is the gross interest rate, w is the average wage, d is dividend, τ is lump-sum tax, L is a lump-sum transfer function, and I is an insurance payments function. Moreover, h_t is agent's health that affects the disutility of working,

$$h_t = \begin{cases} 1 & \text{with probability } \pi \\ \text{draw from } U[0,\eta] & \text{with probability } 1-\pi \end{cases} \quad \text{and} \quad \text{IID},$$

and s_t is individual-specific salary offer that satisfies

$$\ln(s_t) = -\sigma^2/[2(1+\rho)] + \rho \ln s_{t-1} + \varepsilon_t, \quad \{\varepsilon_t\} \overset{\text{IID}}{\sim} N(0,\sigma^2).$$

The Bellman operator related to the household problem is

$$Tv(k,s,h) = \max_{c,n,k'} \left\{ \ln c - \chi(1-h)n + \beta \mathbb{E}_{s',h'|s}\, v(k',s',h') \right\} \tag{36}$$

subject to[17]

$$c, k' \geqslant 0 \quad \text{and} \quad c + k' = (1-\delta+r)k + swn + d - \tau + L(s,h) + I(k,s,h). \tag{37}$$

[17]Households take r, w, d and τ as given.

Regarding parameters above, χ is the disutility from working with the worst possible health, δ is the depreciation rate, and η controls the average utility gap between the healthy and the unhealthy.

Although Oh and Reis (2012) solves the problem via iterating on the standard Bellman operator (36), the computation can be simplified.[18] Consider the plan factorization (W_0, W_1), where

$$\begin{aligned} & W_0 v(s,k') := \mathbb{E}_{s',h'|s}\, v(k',s',h') \\ \text{and} \quad & W_1 g(k,s,h,c,n,k') := \ln c - \chi(1-h)n + \beta g(s,k'). \end{aligned}$$

Then the refactored Bellman operator $S = W_0 M W_1$ is

$$Sg(s,k') = \mathbb{E}_{s',h'|s} \max_{c',n',k''} \left\{ \ln c' - \chi(1-h')n' + \beta g(s',k'') \right\} \tag{38}$$

subject to

$$c', k'' \geqslant 0 \quad \text{and} \quad c' + k'' = (1-\delta+r)k' + s'wn' + d - \tau + L(s',h') + I(k',s',h').$$

Alternatively, we can factorize T by letting

$$\begin{aligned} & W_0 v(k,s) := \mathbb{E}_h v(k,s,h) \\ \text{and} \quad & W_1 g(k,s,h,c,n,k') := \ln c - \chi(1-h)n + \beta \mathbb{E}_{s'|s}\, g(k',s'). \end{aligned}$$

In this case, the refactored Bellman operator $S = W_0 M W_1$ is

$$Sg(k,s) = \mathbb{E}_h \max_{c,n,k'} \left\{ \ln c - \chi(1-h)n + \beta \mathbb{E}_{s'|s}\, g(k',s') \right\} \tag{39}$$

subject to (37). Computations via (38) or (39) are both two-dimensional, while that via (36) is three-dimensional.

5.4. **Consumer Bankruptcy.** Livshits et al. (2007) conducts a quantitative analysis of different consumer bankruptcy rules in an overlapping generation economy. The age j household's income is $z_j \eta_j \bar{e}_j$, where $\bar{e}_j$ is a deterministic labor endowment, and z_j and η_j are respectively the persistent and transitory components of productivity. Let $q_j(d_{j+1}, z_j)$ be the interest rate for age j borrowers, where d_{j+1} is their debt level. Households face an expense shock $\kappa_j \geqslant 0$ that increases their debt (decreases their savings) by the same amount. Moreover, $\{z_j\}$ is Markov, $\{\eta_j\}$ and $\{\kappa_j\}$ are IID, and all shocks are mutually independent.

[18] In the online appendix, Oh and Reis (2012) points out that it is possible to reduce the dimension of the state space from 3 to 2 by *re-defining variables*, which, however, did not speed up computation. We presume that the method mentioned by Oh and Reis is different from our method below.

Beginning each period, households realize productivity and expense shocks, and then decide whether to file for bankruptcy or not. If households do not declare bankruptcy, then they choose their current consumption and next period asset holdings. Households that file for bankruptcy are unable to save in the current period, a fraction γ is deducted from their earnings, and they consume all the remaining. The value of repaying one's debts, v^R, satisfies the Bellman equation

$$v_j^R(d,z,\eta,\kappa) = \max_{c,d'} \left[u(c/n) + \beta \mathbb{E}_{z',\eta',\kappa'|z} \max \left\{ v_{j+1}^R(d',z',\eta',\kappa'),\, v_{j+1}^D(z',\eta') \right\} \right]$$

subject to $c + d + \kappa \leqslant \bar{e}z\eta + q_j(d',z)d'$, where β is discount factor, c is total consumption, n is household size, u is utility function, and v^D is the value of declaring bankruptcy, which satisfies

$$v_j^D(z,\eta) = u(c/n) + \beta \mathbb{E}_{z',\eta',\kappa'|z} \max \left\{ v_{j+1}^R(0,z',\eta',\kappa'),\, v_{j+1}^E(z',\eta',\kappa') \right\}$$

subject to $c = (1-\gamma)\bar{e}z\eta$. In particular, bankruptcy cannot be declared two periods in a row. If a bankrupt household defaults on expense debt, a fraction γ of the household's current income is deducted and its debt is rolled over at a fixed interest rate $\bar{r}$. The value of defaulting on expense debt, denoted by v^E, satisfies

$$v_j^E(z,\eta,\kappa) = u(c/n) + \beta \mathbb{E}_{z',\eta',\kappa'|z} \max \left\{ v_{j+1}^R(d',z',\eta',\kappa'),\, v_{j+1}^D(z',\eta') \right\}$$

subject to $c = (1-\gamma)\bar{e}z\eta$ and $d' = (\kappa - \gamma\bar{e}z\eta)(1+\bar{r})$.

Hence, the Bellman operator T is defined at $(v_{j+1}^R, v_{j+1}^D, v_{j+1}^E)$ by

$$T(v_{j+1}^R, v_{j+1}^D, v_{j+1}^E)(d,z,\eta,\kappa) = \left(v_j^R(d,z,\eta,\kappa),\, v_j^D(z,\eta),\, v_j^E(z,\eta,\kappa) \right).$$

The computation is thus four-dimensional. However, the problem can be simplified via plan factorization. We factorize T as $T = MW_1W_0$. In particular,

$$\begin{aligned} M(h_j^R, h_j^D, h_j^E)(d,z,\eta,\kappa) &:= \left(v_j^R(d,z,\eta,\kappa),\, v_j^D(z,\eta),\, v_j^E(z,\eta,\kappa) \right), \\ W_1(g_j^{RD}, g_j^{RE})(d,z,\eta,\kappa,c,d') &:= \left(h_j^R(d,z,\eta,\kappa,c,d'),\, h_j^D(z,\eta),\, h_j^E(z,\eta,\kappa) \right), \\ \text{and} \quad W_0(v_{j+1}^R, v_{j+1}^D, v_{j+1}^E)(z,d') &:= \left(g_j^{RD}(z,d'),\, g_j^{RE}(z) \right), \end{aligned}$$

where

$$
\begin{aligned}
v_j^R(d, z, \eta, \kappa) &:= \max_{c, d'} h_j^R(d, z, \eta, \kappa, c, d'), \\
v_j^D(z, \eta) &:= h_j^D(z, \eta), \\
v_j^E(z, \eta, \kappa) &:= h_j^E(z, \eta, \kappa), \\
h_j^R(d, z, \eta, \kappa, c, d') &:= u(c/n) + \beta g_j^{RD}(z, d'), \\
h_j^D(z, \eta) &:= u[(1-\gamma)\bar{e} z \eta / n] + \beta g_j^{RE}(z), \\
h_j^E(z, \eta, \kappa) &:= u[(1-\gamma)\bar{e} z \eta / n] + \beta g_j^{RD}[z, (\kappa - \gamma \bar{e} z \eta)(1+\bar{r})], \\
g_j^{RD}(z, d') &:= \mathbb{E}_{z', \eta', \kappa' | z} \max \left\{ v_{j+1}^R(d', z', \eta', \kappa'),\, v_{j+1}^D(z', \eta') \right\}, \\
\text{and} \quad g_j^{RE}(z) &:= \mathbb{E}_{z', \eta', \kappa' | z} \max \left\{ v_{j+1}^R(0, z', \eta', \kappa'),\, v_{j+1}^E(z', \eta', \kappa') \right\}.
\end{aligned}
$$

Then the refactored Bellman operator $S = W_0 M W_1$ is defined at $(g_{j+1}^{RD}, g_{j+1}^{RE})$ by

$$
S(g_{j+1}^{RD}, g_{j+1}^{RE})(z, d') = \left(g_j^{RD}(z, d'),\, g_j^{RE}(z) \right),
$$

where

$$
g_j^{RD}(z, d') = \mathbb{E}_{z', \eta', \kappa' | z} \max \left\{ \max_{c', d''} \left[u(c'/n') + \beta g_{j+1}^{RD}(z', d'') \right],\, u(\tilde{c}/n') + \beta g_{j+1}^{RE}(z') \right\}
$$

subject to

$$
c' + d' + \kappa' \leqslant \bar{e}' z' \eta' + q_{j+1}(d'', z') d'' \quad \text{and} \quad \tilde{c} = (1-\gamma) \bar{e}' z' \eta',
$$

and

$$
g_j^{RE}(z) = \mathbb{E}_{z', \eta', \kappa' | z} \max \left\{ \max_{c', d''} \left[u(c'/n') + \beta g_{j+1}^{RD}(z', d'') \right],\, u(\tilde{c}/n') + \beta g_{j+1}^{RD}(z', \tilde{d}) \right\}
$$

subject to

$$
\begin{aligned}
& c' + \kappa' \leqslant \bar{e}' z' \eta' + q_{j+1}(d'', z') d'', \\
& \tilde{c} = (1-\gamma) \bar{e}' z' \eta' \quad \text{and} \quad \tilde{d} = (\kappa' - \gamma \bar{e}' z' \eta')(1+\bar{r}).
\end{aligned}
$$

While computation via the standard Bellman operator is four-dimensional, computation via the refactored Bellman operator is only two-dimensional.

5.5. **Optimal Default.** Athreya (2008) studies the effect of default, insurance and debt over the life-cycle in an OLG economy. At age j, households receive endowments in the form of income and public transfers. The income follows

$$\ln y_j = \mu_j + z_j + \varepsilon_j, \quad z_j = \gamma z_{j-1} + \eta_j, \quad \{\varepsilon_j\} \overset{\text{IID}}{\sim} N(0, \sigma_\varepsilon^2), \quad \{\eta_j\} \overset{\text{IID}}{\sim} N(0, \sigma_\eta^2),$$

where μ_j, z_j and ε_j are respectively mean log income, and persistent and transitory shocks. Public transfer depends on age j, net assets x_j and income y_j, and guarantees a minimal income level $\underline{\tau}$,

$$\tau_j(x_j, y_j) = \max\{0, \underline{\tau} - [\max(0, x_j) + y_j]\}.$$

Each period, households make the decision to default (if they enter with debt). They are not allowed to issue debt within the period they default. However, they may save and can borrow in subsequent periods. After default decision, the consumption/savings (possibly debt) plan is made. Their value function satisfies

$$v_j(x, z, \varepsilon) = \max\left\{v_j^R(x, z, \varepsilon),\, v_j^D(x, z, \varepsilon)\right\}. \tag{40}$$

Here v_j^R is the value of repaying the debt that comes due at age j, satisfying

$$v_j^R(x, z, \varepsilon) = \max_{c,\, x'}\left\{u(c) + \beta \mathbb{E}_{z', \varepsilon' | z} v_{j+1}(x', z', \varepsilon')\right\}$$

subject to

$$c + x'/R_j(x', z) \leqslant y + \tau_j(x, y) + x, \tag{41}$$

where $u(c)$ is the utility from current period consumption c and $R_j(x', z)$ is the current period interest rate of the (one-period mature) debt.[19]

Moreover, v_j^D denotes the value of eliminating debts via default, satisfying

$$v_j^D(x, z, \varepsilon) = \max_{c,\, x'}\left\{u(c) - \lambda + \beta \mathbb{E}_{z', \varepsilon' | z} v_{j+1}(x', z', \varepsilon')\right\}$$

subject to

$$c + x'/R^f \leqslant y + \tau_j(x, y) \quad \text{and} \quad x' \geqslant 0, \tag{42}$$

where the scalars λ and R^f are respectively the cost of default and annuitized discount rate.

[19]See section 2.5 of Athreya (2008) for how $R_j(x_{j+1}, z_j)$ is endogenously determined.

Hence, the Bellman operator T is defined at v_{j+1} by $Tv_{j+1}(x,z,\varepsilon) = v_j(x,z,\varepsilon)$. Its computation can be simplified by plan factorization. Note that T can be factorized by $T = MW_1W_0$. In particular,

$$M(h_j^R, h_j^D)(x,z,\varepsilon) := \max\left\{h_j^R(x,z,\varepsilon),\, h_j^D(x,z,\varepsilon)\right\},$$

$$W_1 g_j(x,z,\varepsilon) := \left(h_j^R(x,z,\varepsilon),\, h_j^D(x,z,\varepsilon)\right),$$

$$\text{and} \quad W_0 v_{j+1}(x',z) := \mathbb{E}_{z',\varepsilon'|z}\, v_{j+1}(x',z',\varepsilon'),$$

where

$$h_j^R(x,z,\varepsilon) := \max_{c,x'}\left\{u(c) + \beta g_j(x',z)\right\} \quad \text{s.t.} \quad (41)$$

$$\text{and} \quad h_j^D(x,z,\varepsilon) := \max_{c,x'}\left\{u(c) - \lambda + \beta g_j(x',z)\right\} \quad \text{s.t.} \quad (42).$$

Then the refactored Bellman operator $S = W_0MW_1$ is defined at g_{j+1} by

$$g_j(x',z) := Sg_{j+1}(x',z) = \mathbb{E}_{z',\varepsilon'|z} \max\{h_{j+1}^R(x',z',\varepsilon'),\, h_{j+1}^D(x',z',\varepsilon')\}, \tag{43}$$

where

$$h_{j+1}^R(x',z',\varepsilon') := \max_{c',\,x''}\left\{u(c') + \beta g_{j+1}(x'',z')\right\}$$

$$\text{and} \quad h_{j+1}^D(x',z',\varepsilon') := \max_{c',\,x''}\left\{u(c) - \lambda + \beta g_{j+1}(x'',z')\right\},$$

subject respectively to

$$c' + x''/R_{j+1}(x'',z') \leqslant y' + \tau_{j+1}(x',y') + x' \quad \text{and}$$

$$c' + x''/R^f \leqslant y' + \tau_{j+1}(x',y') \quad \text{and} \quad x'' \geqslant 0.$$

Alternatively, we can factorize $T = MW_1W_0$ as

$$M(p_j^R, p_j^D)(x,z,\varepsilon) := \max\left\{p_j^R(x,z,\varepsilon),\, p_j^D(x,z,\varepsilon)\right\},$$

$$W_1 q_{j+1}(x,z,\varepsilon) := \left(p_j^R(x,z,\varepsilon),\, p_j^D(x,z,\varepsilon)\right) \quad \text{and} \quad W_0 v_j(x,z) := \mathbb{E}_{\varepsilon} v_j(x,z,\varepsilon),$$

where

$$p_j^R(x,z,\varepsilon) := \max_{c,x'}\left\{u(c') + \beta \mathbb{E}_{z'|z} q_{j+1}(x',z')\right\} \quad \text{s.t.} \quad (41)$$

$$\text{and} \quad p_j^D(x,z,\varepsilon) := \max_{c,x'}\left\{u(c) - \lambda + \beta \mathbb{E}_{z'|z} q_{j+1}(x',z')\right\} \quad \text{s.t.} \quad (42).$$

Then the refactored Bellman operator $S = W_0MW_1$ is

$$q_j(x,z) := Sq_{j+1}(x,z) = \mathbb{E}_{\varepsilon} \max\left\{p_j^R(x,z,\varepsilon),\, p_j^D(x,z,\varepsilon)\right\}, \tag{44}$$

where p_j^R and p_j^D are defined as above. Computation via either of the refactored Bellman operators ((43) or (44)) is one-dimensional lower than that via the standard Bellman operator (corresponding to the Bellman equation (40)).

5.6. **Portfolio Choice with Annuities.** In an OLG framework, Pashchenko (2013) studies the portfolio choice problem between bonds and annuities. Denote j as the age of an individual, c_j as consumption and k_{j+1} as investment in bonds. Utilities from consumption and bequest are denoted respectively by $u(c_j)$ and $\Gamma(k_j)$. Let $m_j \in \{0, 1\}$ be the agent's health status, which is Markov. A currently alive agent survives to the next period with probability $s_j(m_j, I)$, where I is permanent income. Each period, the agent has to pay a medical cost z_j, satisfying

$$\ln z_j = \mu_j(m_j, I) + \sigma_z \psi_j, \quad \psi_j = \zeta_j + \xi_j \quad \text{and} \quad \zeta_j = \rho \zeta_{j-1} + \varepsilon_j,$$

where $\{\xi_j\} \overset{\text{IID}}{\sim} N(0, \sigma_\xi^2)$ and $\{\varepsilon_j\} \overset{\text{IID}}{\sim} N(0, \sigma_\varepsilon^2)$. An agent without enough resources to pay for medical expenses receives government transfer τ_j, which guarantees a minimum consumption c_{min}. The agent has two investment options: a risk-free bond with return r and an annuity. By paying $q_j \Delta_{j+1}$ today, the agent buys a stream of payments Δ_{j+1} that he will receive each period (if alive). Let n_j denote the total annuity income and $x_j := (I, n_j, k_j)$. The Bellman operator T is

$$\begin{aligned} v_j(x, m, \zeta, \xi) &:= T v_{j+1}(x, m, \zeta, \xi) \\ &= \max_{c, k', \Delta'} \left\{ u(c) + \beta s_j \mathbb{E}_{m', \zeta', \xi' | x, m, \zeta} \, v_{j+1}(x', m', \zeta', \xi') + \beta (1 - s_j) \Gamma(k') \right\} \end{aligned}$$

subject to

$$\begin{aligned} &c + z + k' + q\Delta' = (1 + r)k + n + \tau, \quad n' = \Delta' + n, \quad s_j = s_j(m, I), \\ &\tau = \min\{0, c_{min} - k(1 + r) - n + z\} \quad \text{and} \quad k', \Delta' \geqslant 0. \end{aligned}$$

Consider the plan factorization (W_0, W_1), where

$$W_0 v_j(x, m, \zeta) := \mathbb{E}_{\xi} \, v_j(x, m, \zeta, \xi) \quad \text{and}$$

$$W_1 g_{j+1}(x, m, \zeta, \xi, c, k', \Delta) := u(c) + \beta s_j \mathbb{E}_{m', \zeta' | x, m, \zeta} \, g_{j+1}(x', m', \zeta') + \beta(1 - s_j)\Gamma(k').$$

Then the refactored Bellman operator $S = W_0 M W_1$ is

$$\begin{aligned} g_j(x, m, \zeta) &:= S g_{j+1}(x, m, \zeta) \\ &= \mathbb{E}_{\xi} \max_{c, k', \Delta'} \left\{ u(c) + \beta s_j \mathbb{E}_{m', \zeta' | x, m, \zeta} \, g_{j+1}(x', m', \zeta') + \beta(1 - s_j)\Gamma(k') \right\}. \end{aligned}$$

Computation via the standard Bellman operator T is one-dimensional lower than that via the refactored Bellman operator S.

6. Appendix

In sections 6.1–6.4, we prove all the theoretical results. In section 6.5, we provide counterexamples in support of the theory of section 3.2.

6.1. Preliminaries.

Let E_i be a nonempty set and let τ_i be a mapping from E_i to E_{i+1} for $i = 0, 1, 2$ with addition modulo 3 (a convention we adopt throughout this section). Consider the self-mappings

$$F_0 := \tau_2 \circ \tau_1 \circ \tau_0, \quad F_1 := \tau_0 \circ \tau_2 \circ \tau_1 \quad \text{and} \quad F_2 := \tau_1 \circ \tau_0 \circ \tau_2$$

on E_0, E_1 and E_2 respectively. We then have

$$F_{i+1} \circ \tau_i = \tau_i \circ F_i \quad \text{on } E_i \text{ for } i = 0, 1, 2. \tag{45}$$

Lemma 6.1. *For each $i = 0, 1, 2$,*

(a) if e is a fixed point of F_i in E_i, then $\tau_i e$ is a fixed point of F_{i+1} in E_{i+1}.
(b) $F_{i+1}^n \circ \tau_i = \tau_i \circ F_i^n$ on E_i for all $n \in \mathbb{N}$.

Proof. Regarding part (a), if e is a fixed point of F_i in E_i, then (45) yields $F_{i+1}\tau_i e = \tau_i F_i e = \tau_i e$, so $\tau_i e$ is a fixed point of F_{i+1}. Regarding part (b), fix i in $\{0, 1, 2\}$. By (45), the statement in (b) is true at $n = 1$. Let it also be true at $n - 1$. Then, using (45) again,

$$F_{i+1}^n \circ \tau_i = F_{i+1}^{n-1} \circ F_{i+1} \circ \tau_i = F_{i+1}^{n-1} \circ \tau_i \circ F_i = \tau_i \circ F_i^{n-1} \circ F_i = \tau_i \circ F_i^n.$$

We conclude that (b) holds at every $n \in \mathbb{N}$. □

Lemma 6.2. *If F_i has unique fixed point e_i in E_i for some i in $\{0, 1, 2\}$, then $\tau_i e_i$ is the unique fixed point of F_{i+1} in E_{i+1}.*

Proof. We have already proved all but uniqueness. To see that uniqueness holds, fix $i \in \{0, 1, 2\}$ and suppose that F_i has only one fixed point in E_i, whereas F_{i+1} has two fixed points in E_{i+1}. Denote the fixed points of F_{i+1} by e and f. Applying part (a) of lemma 6.1 twice, we see that $\tau_{i+2}\,\tau_{i+1}\,e$ and $\tau_{i+2}\,\tau_{i+1}\,f$ are both fixed points of $F_{i+3} = F_i$. Since F_i has only one fixed point, we then have

$$\tau_{i+2}\,\tau_{i+1}\,e = \tau_{i+2}\,\tau_{i+1}\,f.$$

Applying τ_i to both sides of the last equality gives $F_{i+1}e = F_{i+1}f$. Since e and f are both fixed points of F_{i+1}, we conclude that $e = f$ and the fixed point is unique. □

6.2. **Multiple Maps.** Consider the same set up as section 6.1, but now, instead of a single map τ_2, suppose instead that we have a family of stage 2 maps $\{\tau_\sigma\}$, each one sending E_2 to E_0, and indexed by $\sigma \in \Sigma$. In addition, suppose that E_i is partially ordered by $\preceq$. The ordering on each set is not necessarily related, although we do not distinguish between them in our notation. Let $\bar{\tau}$ be the pointwise supremum of $\{\tau_\sigma\}$ with respect to $\preceq$. In other words,

$$\bar{\tau} e := \sup_{\sigma \in \Sigma} \tau_\sigma \, e \qquad (e \in E_2). \tag{46}$$

An index $\sigma \in \Sigma$ is called e-*greedy* if $\bar{\tau} e = \tau_\sigma \, e$. We assume that this supremum exists and, moreover, the sup can be replaced with max:

Assumption 6.1. For each e in E_2, there exists at least one e-greedy σ in Σ.

For each $\sigma \in \Sigma$, we consider the self-mappings

$$F_0^\sigma := \tau_\sigma \circ \tau_1 \circ \tau_0, \quad F_1^\sigma := \tau_0 \circ \tau_\sigma \circ \tau_1 \quad \text{and} \quad F_2^\sigma := \tau_1 \circ \tau_0 \circ \tau_\sigma. \tag{47}$$

In addition, define

$$F_0 := \bar{\tau} \circ \tau_1 \circ \tau_0, \quad F_1 := \tau_0 \circ \bar{\tau} \circ \tau_1 \quad \text{and} \quad F_2 := \tau_1 \circ \tau_0 \circ \bar{\tau}. \tag{48}$$

In the following assumption, the “and hence every” statement is due to lemma 6.2.

Assumption 6.2. For each σ in Σ, the map F_i^σ has a unique fixed point e_i^σ in E_i for one, and hence every, $i \in \{0, 1, 2\}$.

When assumption 6.2 holds, we let

$$e_i^* := \sup_{\sigma \in \Sigma} e_i^\sigma \tag{49}$$

for $i \in \{0, 1, 2\}$ whenever the supremum exists.

To simplify the statement and proof of the next proposition, let the one step maps be denoted by R_σ and R respectively. In particular, for fixed σ, the map R_σ equals τ_0 on E_0, τ_1 on E_1 and τ_σ on E_2. The map R is similar but with τ_σ replaced by $\bar{\tau}$. It follows that $F_i^\sigma = R_\sigma^3$ on E_i and $F_i = R^3$ on E_i. Note that, from the definition of $\bar{\tau}$, the map R pointwise dominates R_σ, in the sense that $R \, e \preceq R_\sigma \, e$ for every $e \in E_i$.

Proposition 6.3. *Fix $i \in \{0, 1, 2\}$ and suppose that assumptions 6.1–6.2 hold. If τ_α is isotone for all $\alpha \in \{0, 1\} \cup \Sigma$ and e_i^* is a fixed point of F_i in E_i, then*

$$e_{i+1}^* = R \, e_i^* \tag{50}$$

and e_{i+1}^* is a fixed point of F_{i+1} in E_{i+1}.

Proof. Let the hypotheses in the proposition be true. In view of lemma 6.1, to show that e^*_{i+1} is a fixed point of F_{i+1}, it suffices to show that (50) holds. Pick any $\sigma \in \Sigma$. We have

$$e^\sigma_{i+1} = R_\sigma \, e^\sigma_i \preceq R \, e^\sigma_i \preceq R \, e^*_i,$$

where the equality is due to part (a) of lemma 6.1, the first inequality is due to the fact that R pointwise dominates R_σ and the second follows from the definition of e^*_i. As σ was arbitrary, this proves that $e^*_{i+1} \preceq R \, e^*_i$.

Regarding the reverse inequality, from $F_i \, e^*_i = e^*_i$ and assumption 6.1, which implies the existence of at σ such that $\bar{\tau} = \tau_\sigma$ and hence $F^\sigma_i \, e^*_i = F_i \, e^*_i$, we see that e^*_i is a fixed point of F^σ_i. Therefore, by assumption 6.2, $e^*_i = e^\sigma_i$. As a result, $R \, e^*_i = R_\sigma \, e^\sigma_i = e^\sigma_{i+1}$. But $e^\sigma_{i+1} \preceq e^*_{i+1}$, so $R \, e^*_i \preceq e^*_{i+1}$.

As a partial order, $\preceq$ is by definition antisymmetric, so (50) holds. □

Corollary 6.4. *Assume the conditions of proposition 6.3. If $e^*_i \in E_i$ and $F_i \, e^*_i = e^*_i$ for at least one i, then the same is true for all i, and*

$$e^\sigma_0 = e^*_0 \iff F^\sigma_0 \, e^*_0 = F_0 \, e^*_0 \iff \tau_\sigma \, \tau_1 \, e^*_1 = \bar{\tau} \, \tau_1 \, e^*_1. \tag{51}$$

Proof. If the hypotheses of proposition 6.3 hold, $e^*_i \in E_i$ and $F_i = e^*_i$ for at least one i, then, proposition 6.3 tells us that the same is true for all i, and the fixed points are linked by $e^*_{i+1} = R \, e^*_i$. In this setting we have

$$e^\sigma_0 = e^*_0 \iff F^\sigma_0 \, e^*_0 = F_0 \, e^*_0. \tag{52}$$

Indeed, if $e^\sigma_0 = e^*_0$, then

$$F^\sigma_0 \, e^*_0 = F^\sigma_0 \, e^\sigma_0 = e^\sigma_0 = e^*_0 = F_0 \, e^*_0.$$

Conversely, if $F^\sigma_0 \, e^*_0 = F_0 \, e^*_0$, then, since e^*_0 is in E_0 and, by assumption 6.2, the point e^σ_0 is the only fixed point of F^σ_0 in E_0, we have $e^\sigma_0 = e^*_0$.

We have now verified (52). The last equivalence in (51) follows, since, by the relations $e^*_{i+1} = R \, e^*_i$ and $e^\sigma_{i+1} = R_\sigma \, e^\sigma_i$, we have

$$\tau_\sigma \, \tau_1 \, e^*_1 = \tau_\sigma \, \tau_1 \, \tau_0 \, e^*_0 = F^\sigma_0 \, e^*_0 = F_0 \, e^*_0 = \bar{\tau} \, \tau_1 \, \tau_0 \, e^*_0 = \bar{\tau} \, \tau_1 \, e^*_1. \qquad \square$$

6.3. **Connection to Dynamic Programming.** Now we switch to a dynamic programming setting that builds on the results presented above. When connecting to the results in sections 6.1–6.2, we always take $E_0 = \mathbb{V}$, $E_1 = \mathbb{G}$ and $E_2 = \mathbb{H}$. The partial order $\preceq$ becomes the pointwise partial order. In addition, we set $\tau_0 = W_0$ and $\tau_1 = W_1$. The map τ_2 will vary depending on context.

Proof of proposition 2.2. The claim is immediate from lemmas 6.1–6.2 once we fix $\sigma \in \Sigma$ and set $\tau_2 = M$. □

Proof of proposition 2.3. Similar to the proof of proposition 2.2, this result is immediate from lemmas 6.1–6.2 once we set $\sigma \in \Sigma$ and $\tau_2 = M_\sigma$. □

Proof of theorem 3.2. Assume the conditions of theorem 3.2. Let τ_0 and τ_1 have the identifications given at the start of this section and let $\{\tau_\sigma\}_{\sigma \in \Sigma} = \{M_\sigma\}_{\sigma \in \Sigma}$. Then, fixing $\sigma \in \Sigma$, the following identifications follow:

- $\bar{\tau}$ corresponds to M, since $Mh = \sup_{\sigma \in \Sigma} M_\sigma h$ for all $h \in \mathbb{H}$ by (19).
- $F_0 = \bar{\tau} \circ \tau_1 \circ \tau_0 = M \circ W_1 \circ W_0 = T$, the Bellman operator.
- $F_1 = \tau_0 \circ \bar{\tau} \circ \tau_1 = W_0 \circ M \circ W_1 = S$, the refactored Bellman operator.
- Similarly, F_0^σ and F_1^σ defined in (47) correspond to T_σ and S_σ respectively.
- e_0^σ corresponds to v_σ for each $\sigma \in \Sigma$, and e_0^* corresponds to v^*
- e_1^σ corresponds to g_σ for each $\sigma \in \Sigma$, and e_1^* corresponds to g^*

Note that the conditions of proposition 6.3 are then valid, because W_0 and W_1 are isotone by assumption, M_σ and M are clearly isotone, and assumptions 2.1–2.2, which are held to be true in theorem 3.2, imply assumptions 6.1–6.2.

Now suppose that (a) of theorem 3.2 holds, so that $g^* \in \mathbb{G}$ and $Sg^* = g^*$. We claim that (b) holds, which is to say that $v^* \in \mathbb{V}$ and $Tv^* = v^*$. In the notation of proposition 6.3, we are seeking to show that

$$e_1^* \in E_1 \text{ and } F_1 \, e_1^* = e_1^* \implies e_0^* \in E_0 \text{ and } F_0 \, e_0^* = e_0^*. \tag{53}$$

This follows from proposition 6.3 because if $e_i^* \in E_i$ and $F_i \, e_i^* = e_i^*$ at $i = 1$, then the same is true at $i + 1 = 2$, and hence again at $i + 2 = 0$ with addition modulo 2.

The claim that (b) implies (a) translates to

$$e_0^* \in E_0 \text{ and } F_0 \, e_0^* = e_0^* \implies e_1^* \in E_1 \text{ and } F_1 \, e_1^* = e_1^*, \tag{54}$$

which follows directly from proposition 6.3.

The claim that $g^* = \hat{g}$ under (a)–(b) translates to the claim that $e_1^* = Re_0^*$, which also follows directly from proposition 6.3.

The claim in theorem 3.2 that at least one optimal policy exists under either (a) or (b) follows from the equivalence of (a) and (b) just established combined with theorem 3.1.

Finally, suppose that (a) holds and consider the last claim in theorem 3.2, which can be expressed succinctly as

$$v_\sigma = v^* \iff T_\sigma v^* = T v^* \iff M_\sigma W_1 g^* = M W_1 g^*.$$

In the notation of corollary 6.4, this is precisely (51). Since the conditions of corollary 6.4 hold, the last claim in theorem 3.2 is established. □

6.4. **Remaining Proofs.**

Proof of theorem 3.1. We first show that (a) implies (b). By definition,

$$\sigma^* \text{ is } v^*\text{-greedy} \iff Q(x, \sigma^*(x), v^*) = \max_{a \in \Gamma(x)} Q(x, a, v^*) \iff T_{\sigma^*} v^* = T v^*.$$

Since $v^* = Tv^*$ by (a), $T_{\sigma^*} v^* = T v^*$ implies that $T_{\sigma^*} v^* = v^*$. However, since v_{σ^*} is the unique fixed point of T_{σ^*} in $\mathbb{V}$ by assumption 2.2, we must have $v^* = v_{\sigma^*}$. Hence, $T_{\sigma^*} v^* = T v^*$ if and only if $v^* = v_{\sigma^*}$. We then have

$$\sigma^* \text{ is } v^*\text{-greedy} \iff v^* = v_{\sigma^*} \iff \sigma^* \text{ is optimal},$$

where the second equivalence is due to the definition of optimal policy. Bellman's principle of optimality is verified. Since $v^* \in \mathbb{V}$, assumption 2.1 implies that there exists a v^*-greedy policy $\sigma^* \in \Sigma$. Based on Bellman's principle of optimality, σ^* is optimal, so the set of optimal policies is nonempty. Hence, claim (b) holds.

Next we show that (b) implies (a). Let $\sigma^* \in \Sigma$ be an optimal policy. Then $v_{\sigma^*} = v^*$. Since $v_{\sigma^*} \in \mathbb{V}$ by assumption 2.2, we have $v^* \in \mathbb{V}$. Suppose v^* does not satisfy the Bellman equation. Then there exists $x \in \mathsf{X}$ such that

$$v^*(x) \neq T v^*(x) = \max_{a \in \Gamma(x)} Q(x, a, v^*). \tag{55}$$

Based on Bellman's principle of optimality, σ^* is v^*-greedy, which implies

$$Q(x, \sigma^*(x), v^*) = \max_{a \in \Gamma(x)} Q(x, a, v^*).$$

Since $v_{\sigma^*} = v^*$, the definition of v_{σ^*} and assumption 2.2 imply that

$$v^*(x) = v_{\sigma^*}(x) = Q(x, \sigma^*(x), v_{\sigma^*}) = Q(x, \sigma^*(x), v^*) = \max_{a \in \Gamma(x)} Q(x, a, v^*).$$

However, this is contradicted with (55). Hence, v^* satisfies the Bellman operator and claim (a) holds. This concludes the proof. □

Proof of theorem 3.3. Assume the hypotheses of the theorem. We need only show that (a) holds, since the remaining claims follow directly from these hypotheses, claim (a), the definition of asymptotic stability (for part (b)) and theorem 3.2 (for parts (c) and (d)).

To see that (a) holds, note that, by the stated hypotheses, S has a unique fixed point in $\mathbb{G}$, which we denote below by $\bar{g}$. Our aim is to show that $\bar{g} = g^*$.

First observe that, by existence of greedy policies, there is a $\sigma \in \Sigma$ such that $S_\sigma \bar{g} = S\bar{g} = \bar{g}$. But, by assumption 2.2 and proposition 2.3, S_σ has exactly one fixed point, which is the refactored σ-value function g_σ. Hence $\bar{g} = g_\sigma$. In particular, $\bar{g} \leqslant g^*$.

To see that the reverse inequality holds, pick any $\sigma \in \Sigma$ and note that, by the definition of S, we have $\bar{g} = S\bar{g} \geqslant S_\sigma \bar{g}$. We know that S_σ is isotone on $\mathbb{G}$, since this operator is the composition of three isotone operators (W_0 and W_1 by assumption and M_σ automatically). Hence we can iterate on the last inequality to establish $\bar{g} \geqslant S_\sigma^k \bar{g}$ for all $k \in \mathbb{N}$. Taking limits and using the fact that the partial order is closed, S_σ is asymptotically stable and, as shown above, g_σ is the unique fixed point, we have $\bar{g} \geqslant g_\sigma$. Since σ was chosen arbitrarily, this yields $\bar{g} \geqslant g^*$. That part (a) of theorem 3.3 holds is now established. □

Proof of proposition 3.4. As a closed subset of $b_\kappa \mathsf{F}$ under the $\| \cdot \|_\kappa$-norm metric, the set $\mathbb{G}$ is complete under the same metric and, by (23) and the Banach contraction mapping theorem, each S_σ is asymptotically stable. Moreover, the pointwise partial order $\leqslant$ is closed under this metric. Thus, to verify the conditions of theorem 3.3, we need only show that S is also asymptotically stable under the same metric.

To this end, we first claim that, under the stated assumptions,

$$Sg(x, a) = \max_{\sigma \in \Sigma} S_\sigma g(x, a), \quad \text{for all } (x, a) \in \mathsf{F} \text{ and } g \in \mathbb{G}. \tag{56}$$

To see that (56) holds, fix $g \in \mathbb{G}$ and observe that, by assmption 2.1, we have $M\, W_1\, g = \sup_{\sigma \in \Sigma} M_\sigma\, W_1\, g$ when the supremum is defined in terms of pointwise order. It follows that, for any $\sigma \in \Sigma$ we have $M\, W_1\, g \geqslant M_\sigma\, W_1\, g$ and hence, applying W_0 to both sides and using isotonicity yields $Sg \geqslant S_\sigma g$. Moreover, if σ is a g-greedy policy then by definition we have $M\, W_1\, g = M_\sigma\, W_1\, g$ and applying W_0 to both sides again gives $Sg = S_\sigma g$ at this σ. Hence (56) is valid.

Now fix $g, g' \in \mathbb{G}$ and $(x, a) \in \mathsf{F}$. By (56) and the contraction condition on S_σ in (23), we have

$$
\begin{aligned}
|Sg(x,a) - Sg'(x,a)| &= |\max_{\sigma \in \Sigma} S_\sigma g(x,a) - \max_{\sigma \in \Sigma} S_\sigma g'(x,a)| \\
&\leqslant \max_{\sigma \in \Sigma} |S_\sigma g(x,a) - S_\sigma g'(x,a)|.
\end{aligned}
$$

Therefore

$$
\frac{|Sg(x,a) - Sg'(x,a)|}{\kappa(x,a)} \leqslant \max_{\sigma \in \Sigma} \|S_\sigma g - S_\sigma g'\|_\kappa \leqslant \beta \|g - g'\|_\kappa.
$$

Taking the supremum gives $\|Sg - Sg'\|_\kappa \leqslant \beta \, \|g - g'\|_\kappa$, so S is also asymptotically stable. Thus, all the conditions of theorem 3.3 are verified and its conclusions follow. □

Proof of proposition 3.5. Without loss of generality, we assume that $m_k \equiv m$ for all $k \in \mathbb{N}_0$. Let Σ_k^v and Σ_k^g denote respectively the set of v_k-greedy and g_k-greedy policies. When $k = 0$, by the definition of greedy policies, we have

$$
\Sigma_0^v = \{\sigma \in \Sigma \colon M_\sigma W_1 W_0 v_0 = M W_1 W_0 v_0\} = \{\sigma \in \Sigma \colon M_\sigma W_1 g_0 = M_\sigma W_1 g_0\} = \Sigma_0^g,
$$

i.e., the sets of v_0-greedy and g_0-greedy policies are identical, with $\Sigma_0 = \Sigma_0^v = \Sigma_0^g$. This in turn implies that a policy $\sigma_0 \in \Sigma$ is v_0-greedy if and only if it is g_0-greedy. Moreover, by proposition 2.1, we have

$$
\begin{aligned}
g_1 = S_{\sigma_0}^m g_0 &= W_0 T_{\sigma_0}^{m-1} M_{\sigma_0} W_1 g_0 = W_0 T_{\sigma_0}^{m-1} M_{\sigma_0} W_1 W_0 v_0 \\
&= W_0 T_{\sigma_0}^{m-1} T_{\sigma_0} v_0 = W_0 T_{\sigma_0}^m v_0 = W_0 v_1
\end{aligned}
$$

We have thus verified the related claims for $k = 0$. Suppose these claims hold for arbitrary k. It remains to show that they hold for $k+1$. By the induction hypothesis, $\Sigma_k^v = \Sigma_k^g = \Sigma_k$, a policy $\sigma_k \in \Sigma$ is v_k-greedy if and only if it is g_k-greedy, and $g_{k+1} = W_0 v_{k+1}$. By the definition of greedy policies, we have

$$
\begin{aligned}
\Sigma_{k+1}^v &= \{\sigma \in \Sigma \colon M_\sigma W_1 W_0 v_{k+1} = M W_1 W_0 v_{k+1}\} \\
&= \{\sigma \in \Sigma \colon M_\sigma W_1 g_{k+1} = M W_1 g_{k+1}\} = \Sigma_{k+1}^g = \Sigma_{k+1}
\end{aligned}
$$

and a policy $\sigma_{k+1} \in \Sigma$ is v_{k+1}-greedy if and only if it is g_{k+1}-greedy. Moreover, proposition 2.1 implies that

$$
\begin{aligned}
g_{k+2} = S_{\sigma_{k+1}}^m g_{k+1} &= W_0 T_{\sigma_{k+1}}^{m-1} M_{\sigma_{k+1}} W_1 g_{k+1} \\
&= W_0 T_{\sigma_{k+1}}^{m-1} M_{\sigma_{k+1}} W_1 W_0 v_{k+1} = W_0 T_{\sigma_{k+1}}^m v_{k+1} = W_0 v_{k+2}.
\end{aligned}
$$

Hence, the related claims of the proposition hold for $k+1$, completing the proof. □

Proof of theorem 3.6. Without loss of generality, we assume $m_k \equiv m$ for all $k \in \mathbb{N}_0$. Let $\{\omega_k\}$ be defined by $\omega_0 := g_0$ and $\omega_k := S\,\omega_{k-1}$. We show by induction that

$$\omega_k \leqslant g_k \leqslant g^* \quad \text{and} \quad g_k \leqslant Sg_k \quad \text{for all } k \in \mathbb{N}_0. \tag{57}$$

Note that $g_0 \leqslant Sg_0$ by assumption. Since S is isotone, this implies that $g_0 \leqslant S^t g_0$ for all $t \in \mathbb{N}_0$. Letting $t \to \infty$, theorem 3.3 implies that $g_0 \leqslant g^*$. Since in addition $\omega_0 = g_0$, (57) is satisfied for $k = 0$. Suppose this claim holds for arbitrary $k \in \mathbb{N}_0$. Then, since S and S_{σ_k} are isotone,

$$g_{k+1} = S_{\sigma_k}^m g_k \leqslant SS_{\sigma_k}^{m-1} g_k \leqslant SS_{\sigma_k}^{m-1} Sg_k = SS_{\sigma_k}^m g_k = Sg_{k+1},$$

where the first inequality holds by the definition of S, the second inequality holds since $g_k \leqslant Sg_k$ (the induction hypothesis), and the second equality is due to $\sigma_k \in \Sigma_k$. Hence, $g_{k+1} \leqslant S^t g_{k+1}$ for all $t \in \mathbb{N}_0$. Letting $t \to \infty$ yields $g_{k+1} \leqslant g^*$.

Similarly, since $g_k \leqslant Sg_k = S_{\sigma_k} g_k$, we have $g_k \leqslant S_{\sigma_k}^{m-1} g_k$ and thus

$$\omega_{k+1} = S\omega_k \leqslant Sg_k = S_{\sigma_k} g_k \leqslant S_{\sigma_k}^m g_k = g_{k+1}.$$

Hence, (57) holds by induction. The isotonicity of S_{σ_k} implies that

$$g_k \leqslant Sg_k = S_{\sigma_k} g_k \leqslant \cdots \leqslant S_{\sigma_k}^m g_k = g_{k+1}$$

for all $k \in \mathbb{N}_0$. Hence, $\{g_k\}$ is increasing in k. Moreover, since $\omega_k \to g^*$ by theorem 3.3, (57) implies that $g_k \to g^*$ as $k \to \infty$. $\square$

Proof of lemma 4.1. Fix $g \in \mathbb{G}$. Since g is bounded below, $h_g(0) = r(0,0) + \beta g(0) \in \mathbb{R} \cup \{-\infty\}$ and h_g is well defined at $x = 0$. For all $x > 0$, since g is continuous, r is continuous on D_0, and $r(x, a) \to -\infty$ as $a \to x$ if r is not continuous on D/D_0, the maximum in the definition of h_g can be attained at some $x_0 \in [0, x]$ and h_g is well defined at x. Hence, h_g is well defined on $\mathbb{R}_+$. Regarding monotonicity, let $x_1, x_2 \in \mathbb{R}_+$ with $x_1 < x_2$. By the monotonicity of r (w.r.t x), we have

$$h_g(x_1) \leqslant \max_{a \in [0,x_1]} \{r(x_2, a) + \beta g(a)\} \leqslant \max_{a \in [0,x_2]} \{r(x_2, a) + \beta g(a)\} = h_g(x_2).$$

Hence, claim (a) holds. Claims (b)–(d) follow from the theorem of maximum (see, e.g., theorem 3.6 of Stokey et al. (1989)), adjusted to accommodate the case of possibly negative infinity valued objective functions. $\square$

Proof of theorem 4.2. To prove the desired results, we only need to verify that the assumptions of proposition 3.4 hold in the current context. Note that W_0 and W_1 are isotone by construction. Moreover, one can show that $\mathbb{G}$ is a closed subset of $b_\kappa \mathsf{F}$.

To verify assumption 2.1, it suffices to show the existence of g-greedy policies. Fix $g \in \mathbb{G}$. By lemma 4.1 and the measurable selection theorem (see, e.g., theorem 7.6 of Stokey et al. (1989)), there exists a measurable map $\sigma : \mathsf{X} \to \mathsf{A}$ such that $\sigma(x) \in \Gamma(x)$ and $MW_1 g = M_\sigma W_1 g$. Hence, $Sg = S_\sigma g$. To show that σ is g-greedy, it remains to verify $S_\sigma g \in \mathbb{G}$ (which guarantees $\sigma \in \Sigma$). We first show that $S_\sigma g$ is Borel measurable. The monotonicity of h_g established in lemma 4.1 implies that h_g is Borel measurable. Since f is Borel measurable, $h_g \circ f \colon \mathbb{R}_+^2 \to \mathbb{R} \cup \{-\infty\}$ is also Borel measurable. Assumption 4.1-(f) then implies that

$$\mathbb{R}_+ \ni a \mapsto S_\sigma g(a) = Sg(a) = (\mathcal{R} \circ h_g \circ f)(a, \xi) \in \mathbb{R}$$

is Borel measurable.

Next, we show that $S_\sigma g$ is bounded below. Since g is bounded below, there exists $M \in \mathbb{R}$ such that $g \geqslant M$. Then, by the monotonicity of r and f,

$$\begin{aligned}(h_g \circ f)(a, z) &\geqslant \max_{a' \in [0, f(a,z)]} r\left(f(a, z), a'\right) + \beta M \\ &= r(f(a, z), 0) + \beta M \geqslant r(f(0, z), 0) + \beta M.\end{aligned}$$

Hence, by the monotonicity of $\mathcal{R}$ and assumption 4.1-(g),

$$\begin{aligned}S_\sigma g(a) &= Sg(a) = (\mathcal{R} \circ h_g \circ f)(a, \xi) \\ &\geqslant \mathcal{R}\left(r(f(0, \xi), 0) + \beta M\right) \geqslant \mathcal{R} r(f(0, \xi), 0) + \beta M > -\infty\end{aligned}$$

for all $a \in \mathbb{R}_+$. Hence, $S_\sigma g$ is bounded below.

Our next job is to show that $S_\sigma g$ is κ-bounded above. By the definition of $\mathcal{R}$, conditions (a), (c) and (e) of assumption 4.1, and the properties of g, there exists $K \in \mathbb{R}_+$ such that

$$\begin{aligned}S_\sigma g(a) &= \mathcal{R}\{r(f(a, \xi), \sigma[f(a, \xi)]) + \beta g(\sigma[f(z, \xi)])\} \\ &\leqslant \mathcal{R}\{r(f(a, \xi), 0) + \beta g[f(z, \xi)]\} \\ &\leqslant \mathcal{R}\{(d + \beta K)\kappa[f(a, \xi)]\} \leqslant \mathbb{E}\left\{(d + \beta K)\kappa[f(a, \xi)]\right\} \\ &= (d + \beta K) \int_{\mathbb{R}_+} \kappa[f(a, z)] \mu(\mathrm{d}z) \leqslant (d + \beta K) \alpha\, \kappa(a)\end{aligned}$$

for all $a \in \mathbb{R}_+$. Hence, $\sup_{a \in \mathbb{R}_+} [S_\sigma g(a) / \kappa(a)] < \infty$.

Next, we show that $S_\sigma g$ is increasing. For all $a_1, a_2 \in \mathbb{R}_+$ with $a_1 < a_2$, the monotonicity of f (w.r.t a) implies that $f(a_1, z) \leqslant f(a_2, z)$. By lemma 4.1, $(h_g \circ f)(a_1, z) \leqslant (h_g \circ f)(a_2, z)$. The monotonicity of $\mathcal{R}$ then implies that $S_\sigma g(a_1) = Sg(a_1) = (\mathcal{R} \circ h_g \circ f)(a_1, \xi) \leqslant (\mathcal{R} \circ h_g \circ f)(a_2, \xi) = Sg(a_2) = S_\sigma g(a_2)$.

We then show that $S_\sigma g$ is continuous. By assumption 4.1-(d), $a \mapsto f(a, \xi)$ is strictly positive and continuous μ-almost surely. Lemma 4.1 then implies that $a \mapsto (h_g \circ f)(a, \xi)$ is continuous μ-almost surely. Moreover, the above analysis implies that

$$r(f(0,z),0) + \beta M \leqslant (h_g \circ f)(a,z) \leqslant (d + \beta K)\kappa[f(a,z)].$$

By assumption 4.1-(h), $a \mapsto S_\sigma g(a) = Sg(a) = (\mathcal{R} \circ h_g \circ f)(a, \xi)$ is continuous. Now we have verified that $S_\sigma g \in \mathbb{G}$. Hence, assumption 2.1 holds.

Finally, we need to verify assumption 2.2 and (23). Note that the latter implies the former. Fix $\sigma \in \Sigma$ and $g_1, g_2 \in \mathbb{G}$. Then, by conditions (a), (b), (d) and (e) of assumption 4.1, we have

$$\begin{aligned}
S_\sigma g_1(a) - S_\sigma g_2(a) &= \mathcal{R}\{r(f(a,\xi), \sigma[f(a,\xi)]) + \beta g_1(\sigma[f(a,\xi)])\} \\
&\qquad - \mathcal{R}\{r(f(a,\xi), \sigma[f(a,\xi)]) + \beta g_2(\sigma[f(a,\xi)]\} \\
&\leqslant \mathcal{R}[\beta g_1(\sigma[f(a,\xi)]) - \beta g_2(\sigma[f(a,\xi)])] \\
&\leqslant \mathcal{R}\left(\beta \|g_1 - g_2\|_\kappa \, \kappa[f(a,\xi)]\right) \leqslant \mathbb{E}\left(\beta \|g_1 - g_2\|_\kappa \, \kappa[f(a,\xi)]\right) \\
&= \beta \|g_1 - g_2\|_\kappa \int_{\mathbb{R}_+} \kappa[f(a,z)] \mu(\mathrm{d}z) \leqslant \alpha\beta \|g_1 - g_2\|_\kappa \, \kappa(a)
\end{aligned}$$

for all $a \in \mathbb{R}_+$. Switching the roles of g_1 and g_2, we get

$$\|S_\sigma g_1 - S_\sigma g_2\|_\kappa \leqslant \alpha\beta \|g_1 - g_2\|_\kappa.$$

We have now verified all the assumptions of proposition 3.4. Hence, all the statements of theorem 4.2 hold. $\square$

6.5. **Counterexamples.** Counterexamples showing that isotonicity of W_0 and W_1 cannot be dropped from theorems 3.2 are provided in this section.

[add other theorems]

6.5.1. *Counterexample 1.* Here we exhibit a dynamic program and value transformation under which $Tv^* = v^*$ and yet $Sg^* \neq g^*$. In addition, the functions g^* and $\hat{g}$ do not agree. The example involves risk sensitive preferences.

Let $\mathsf{X} := \{1, 2\}$, $\mathsf{A} := \{0, 1\}$, $\mathbb{V} := \mathbb{R}^{\mathsf{X}}$ and $\Sigma := \{\sigma_1, \sigma_2, \sigma_3, \sigma_4\}$, where

$$\begin{aligned}
&\sigma_1(1) = 0, \quad \sigma_1(2) = 0, \quad \sigma_2(1) = 0, \quad \sigma_2(2) = 1, \\
&\sigma_3(1) = 1, \quad \sigma_3(2) = 1, \quad \sigma_4(1) = 1, \quad \sigma_4(2) = 0.
\end{aligned}$$

In state x, choosing action a provides the agent with an immediate reward $x - a$. If $a = 0$, then the next period state $x' = 1$, while if $a = 1$, the next period state $x' = 2$. The Bellman operator T is

$$\begin{aligned} Tv(x) &= \max_{a \in \{0,1\}} \left\{ x - a - \frac{\beta}{\gamma} \ln \mathbb{E}_a \exp[-\gamma v(x')] \right\} \\ &= \max \{ x + \beta v(1), \, x - 1 + \beta v(2) \} . \end{aligned}$$

Suppose that we set

$$W_0 \, v(a) := \mathbb{E}_a \exp[-\gamma v(x')] \quad \text{and} \quad W_1 \, g(x, a) := x - a - \frac{\beta}{\gamma} \ln g(a).$$

The refactored Bellman operator is therefore

$$Sg(a) = \mathbb{E}_a \exp \left[-\gamma \max_{a' \in \{0,1\}} \left\{ x' - a' - \frac{\beta}{\gamma} \ln g(a') \right\} \right] .$$

Note that neither W_1 nor W_0 is isotone. In the following, we assume that $\beta \in (0, 1)$.

Lemma 6.5. *The σ-value functions are given by*

$$\begin{aligned} & v_{\sigma_1}(1) = 1/(1-\beta), \quad v_{\sigma_1}(2) = (2-\beta)/(1-\beta), \\ & v_{\sigma_2}(1) = 1/(1-\beta), \quad v_{\sigma_2}(2) = 1/(1-\beta), \\ & v_{\sigma_3}(1) = \beta/(1-\beta), \quad v_{\sigma_3}(2) = 1/(1-\beta), \\ & v_{\sigma_4}(1) = (2\beta)/(1-\beta^2), \quad v_{\sigma_4}(2) = 2/(1-\beta^2). \end{aligned}$$

Based on lemma 6.5, the value function v^* satisfies

$$\begin{aligned} v^*(1) &= \max_{\sigma_i} v_{\sigma_i}(1) = v_{\sigma_1}(1) = v_{\sigma_2}(1) = 1/(1-\beta) \\ v^*(2) &= \max_{\sigma_i} v_{\sigma_i}(2) = v_{\sigma_1}(2) = (2-\beta)/(1-\beta). \end{aligned}$$

On the other hand,

$$\begin{aligned} Tv^*(1) &= \max \left\{ 1 - \frac{\beta}{\gamma} \ln \mathbb{E}_0 \exp[-\gamma v^*(x')], \, -\frac{\beta}{\gamma} \ln \mathbb{E}_1 \exp[-\gamma v^*(x')] \right\} \\ &= \max\{1 + \beta v^*(1), \, \beta v^*(2)\} = 1/(1-\beta) = v^*(1) \end{aligned}$$

$$\begin{aligned} \text{and} \quad Tv^*(2) &= \max \left\{ 2 - \frac{\beta}{\gamma} \ln \mathbb{E}_0 \exp[-\gamma v^*(x')], \, 1 - \frac{\beta}{\gamma} \ln \mathbb{E}_1 \exp[-\gamma v^*(x')] \right\} \\ &= \max\{2 + \beta v^*(1), \, 1 + \beta v^*(2)\} = (2-\beta)/(1-\beta) = v^*(2). \end{aligned}$$

Hence, $v^* = Tv^*$ as claimed.

Lemma 6.6. *The refactored σ-value functions are given by*

$$
\begin{aligned}
g_{\sigma_1}(0) &= \exp[-\gamma/(1-\beta)], \quad g_{\sigma_1}(1) = \exp[-\gamma(2-\beta)/(1-\beta)], \\
g_{\sigma_2}(0) &= \exp[-\gamma/(1-\beta)], \quad g_{\sigma_2}(1) = \exp[-\gamma/(1-\beta)], \\
g_{\sigma_3}(0) &= \exp[-\gamma\beta/(1-\beta)], \quad g_{\sigma_3}(1) = \exp[-\gamma/(1-\beta)], \\
g_{\sigma_4}(0) &= \exp[-2\gamma\beta/(1-\beta^2)], \quad g_{\sigma_4}(1) = \exp[-2\gamma/(1-\beta^2)].
\end{aligned}
$$

Based on lemma 6.6, the refactored value function g^* satisfies

$$
\begin{aligned}
& g^*(0) = \max_{\sigma_i} g_{\sigma_i}(0) = g_{\sigma_3}(0) = \exp[-\gamma\beta/(1-\beta)] \\
\text{and} \quad & g^*(1) = \max_{\sigma_i} g_{\sigma_i}(0) = g_{\sigma_2}(1) = g_{\sigma_3}(1) = \exp[-\gamma/(1-\beta)].
\end{aligned}
$$

Since $\hat{g}(0) = W_0 v^*(0) = \exp[-\gamma v^*(1)] = \exp[-\gamma/(1-\beta)] \neq g^*(0)$, we know that $g^* \neq \hat{g}$. Moreover,

$$
\begin{aligned}
Sg^*(0) &= \mathbb{E}_0 \exp\left(-\gamma \max_{a' \in \{0,1\}} \left\{x' - a' - \frac{\beta}{\gamma} \ln g^*(a')\right\}\right) \\
&= \exp\left(-\gamma \max\left\{1 - \frac{\beta}{\gamma} \ln g^*(0), -\frac{\beta}{\gamma} \ln g^*(1)\right\}\right) \\
&= \exp[-\gamma(1-\beta+\beta^2)/(1-\beta^2)] \neq g^*(0).
\end{aligned}
$$

Hence, $g^* \neq Sg^*$ and the refactored value function is not a fixed point of the refactored Bellman operator.

Proof of lemma 6.5. Regarding v_{σ_1}, by definition,

$$
\begin{aligned}
& v_{\sigma_1}(1) = 1 - \sigma_1(1) - \frac{\beta}{\gamma} \ln \mathbb{E}_{\sigma_1(1)} \exp[-\gamma v_{\sigma_1}(x')] = 1 + \beta v_{\sigma_1}(1) \\
\text{and} \quad & v_{\sigma_1}(2) = 2 - \sigma_1(2) - \frac{\beta}{\gamma} \ln \mathbb{E}_{\sigma_1(2)} \exp[-\gamma v_{\sigma_1}(x')] = 2 + \beta v_{\sigma_1}(1).
\end{aligned}
$$

Hence, $v_{\sigma_1}(1) = 1/(1-\beta)$ and $v_{\sigma_1}(2) = (2-\beta)/(1-\beta)$. Regarding v_{σ_2}, by definition,

$$
\begin{aligned}
& v_{\sigma_2}(1) = 1 - \sigma_2(1) - \frac{\beta}{\gamma} \ln \mathbb{E}_{\sigma_2(1)} \exp[-\gamma v_{\sigma_2}(x')] = 1 + \beta v_{\sigma_2}(1) \\
\text{and} \quad & v_{\sigma_2}(2) = 2 - \sigma_2(2) - \frac{\beta}{\gamma} \ln \mathbb{E}_{\sigma_2(2)} \exp[-\gamma v_{\sigma_2}(x')] = 1 + \beta v_{\sigma_2}(2).
\end{aligned}
$$

Hence, $v_{\sigma_2}(1) = v_{\sigma_2}(2) = 1/(1-\beta)$.

Regarding v_{σ_3}, by definition,

$$v_{\sigma_3}(1) = 1 - \sigma_3(1) - \frac{\beta}{\gamma} \ln \mathbb{E}_{\sigma_3(1)} \exp[-\gamma v_{\sigma_3}(x')] = \beta v_{\sigma_3}(2)$$

$$\text{and} \quad v_{\sigma_3}(2) = 2 - \sigma_3(2) - \frac{\beta}{\gamma} \ln \mathbb{E}_{\sigma_3(2)} \exp[-\gamma v_{\sigma_3}(x')] = 1 + \beta v_{\sigma_3}(2).$$

Hence, $v_{\sigma_3}(1) = \beta/(1-\beta)$ and $v_{\sigma_3}(2) = 1/(1-\beta)$.

Regarding v_{σ_4}, by definition,

$$v_{\sigma_4}(1) = 1 - \sigma_4(1) - \frac{\beta}{\gamma} \ln \mathbb{E}_{\sigma_4(1)} \exp[-\gamma v_{\sigma_4}(x')] = \beta v_{\sigma_4}(2)$$

$$\text{and} \quad v_{\sigma_4}(2) = 2 - \sigma_4(2) - \frac{\beta}{\gamma} \ln \mathbb{E}_{\sigma_4(2)} \exp[-\gamma v_{\sigma_4}(x')] = 2 + \beta v_{\sigma_4}(1).$$

Hence, $v_{\sigma_4}(1) = 2\beta/(1-\beta^2)$ and $v_{\sigma_4}(2) = 2/(1-\beta^2)$. □

Proof of lemma 6.6. Regarding g_{σ_1}, by definition,

$$\begin{aligned} g_{\sigma_1}(0) &= \mathbb{E}_0 \exp\left(-\gamma\left\{x' - \sigma_1(x') - \frac{\beta}{\gamma} \ln g_{\sigma_1}[\sigma_1(x')]\right\}\right) \\ &= \exp\left(-\gamma\left[1 - \frac{\beta}{\gamma} \ln g_{\sigma_1}(0)\right]\right) = \exp(-\gamma) g_{\sigma_1}(0)^\beta \end{aligned}$$

$$\begin{aligned} \text{and} \quad g_{\sigma_1}(1) &= \mathbb{E}_1 \exp\left(-\gamma\left\{x' - \sigma_1(x') - \frac{\beta}{\gamma} \ln g_{\sigma_1}[\sigma_1(x')]\right\}\right) \\ &= \exp\left(-\gamma\left[2 - \frac{\beta}{\gamma} \ln g_{\sigma_1}(0)\right]\right) = \exp(-2\gamma) g_{\sigma_1}(0)^\beta. \end{aligned}$$

Hence, $g_{\sigma_1}(0) = \exp[-\gamma/(1-\beta)]$ and $g_{\sigma_1}(1) = \exp[-\gamma(2-\beta)/(1-\beta)]$.

Regarding g_{σ_2}, by definition,

$$\begin{aligned} g_{\sigma_2}(0) &= \mathbb{E}_0 \exp\left(-\gamma\left\{x' - \sigma_2(x') - \frac{\beta}{\gamma} \ln g_{\sigma_2}[\sigma_2(x')]\right\}\right) \\ &= \exp\left(-\gamma\left[1 - \frac{\beta}{\gamma} \ln g_{\sigma_2}(0)\right]\right) = \exp(-\gamma) g_{\sigma_2}(0)^\beta \end{aligned}$$

$$\begin{aligned} \text{and} \quad g_{\sigma_2}(1) &= \mathbb{E}_1 \exp\left(-\gamma\left\{x' - \sigma_2(x') - \frac{\beta}{\gamma} \ln g_{\sigma_2}[\sigma_2(x')]\right\}\right) \\ &= \exp\left(-\gamma\left[1 - \frac{\beta}{\gamma} \ln g_{\sigma_2}(1)\right]\right) = \exp(-\gamma) g_{\sigma_2}(1)^\beta. \end{aligned}$$

Hence, $g_{\sigma_2}(0) = g_{\sigma_2}(1) = \exp[-\gamma/(1-\beta)]$.

Regarding g_{σ_3}, by definition,

$$g_{\sigma_3}(0) = \mathbb{E}_0 \exp\left(-\gamma\left\{x' - \sigma_3(x') - \frac{\beta}{\gamma}\ln g_{\sigma_3}[\sigma_3(x')]\right\}\right)$$
$$= \exp\left(-\gamma\left[-\frac{\beta}{\gamma}\ln g_{\sigma_3}(1)\right]\right) = g_{\sigma_3}(1)^{\beta}.$$

$$\text{and}\quad g_{\sigma_3}(1) = \mathbb{E}_1 \exp\left(-\gamma\left\{x' - \sigma_3(x') - \frac{\beta}{\gamma}\ln g_{\sigma_3}[\sigma_3(x')]\right\}\right)$$
$$= \exp\left(-\gamma\left[1-\frac{\beta}{\gamma}\ln g_{\sigma_3}(1)\right]\right) = \exp(-\gamma) g_{\sigma_3}(1)^{\beta}.$$

Hence, $g_{\sigma_2}(0) = \exp[-\gamma\beta/(1-\beta)]$ and $g_{\sigma_3}(1) = \exp[-\gamma/(1-\beta)]$.

Regarding g_{σ_4}, by definition,

$$g_{\sigma_4}(0) = \mathbb{E}_0 \exp\left(-\gamma\left\{x' - \sigma_4(x') - \frac{\beta}{\gamma}\ln g_{\sigma_4}[\sigma_4(x')]\right\}\right)$$
$$= \exp\left(-\gamma\left[-\frac{\beta}{\gamma}\ln g_{\sigma_4}(1)\right]\right) = g_{\sigma_4}(1)^{\beta}.$$

$$\text{and}\quad g_{\sigma_4}(1) = \mathbb{E}_1 \exp\left(-\gamma\left\{x' - \sigma_4(x') - \frac{\beta}{\gamma}\ln g_{\sigma_4}[\sigma_4(x')]\right\}\right)$$
$$= \exp\left(-\gamma\left[2-\frac{\beta}{\gamma}\ln g_{\sigma_4}(0)\right]\right) = \exp(-2\gamma) g_{\sigma_4}(0)^{\beta}.$$

Hence, $g_{\sigma_2}(0) = \exp[-2\gamma\beta/(1-\beta^2)]$ and $g_{\sigma_4}(1) = \exp[-2\gamma/(1-\beta^2)]$. □

6.5.2. *Counterexample 2.* We now exhibit a dynamic program and value transformation under which $Sg^* = g^*$ but $Tv^* \neq v^*$. Moreover, the functions g and $\hat{g}$ do not agree.

The set up is the same as the previous counterexample in section 6.5.1, except that we assume $\beta > 1$. In this case, lemmas 6.5–6.6 still hold. Moreover, as in section 6.5.1, the value function and refactored value function satisfy

$$v^*(1) = 1/(1-\beta), \quad v^*(2) = (2-\beta)/(1-\beta),$$
$$g^*(0) = \exp[-\gamma\beta/(1-\beta)], \quad g^*(1) = \exp[-\gamma/(1-\beta)].$$

Recall that in section 6.5.1 we have also shown $g^* \neq \hat{g}$. Moreover, since

$$\begin{aligned} Sg^*(0) &= \mathbb{E}_0 \exp\left(-\gamma \max_{a' \in \{0,1\}} \left\{x' - a' - \frac{\beta}{\gamma} \ln g^*(a')\right\}\right) \\ &= \exp\left(-\gamma \max\left\{1 - \frac{\beta}{\gamma} \ln g^*(0), -\frac{\beta}{\gamma} \ln g^*(1)\right\}\right) \\ &= \exp[-\gamma\beta/(1-\beta)] = g^*(0) \end{aligned}$$

$$\text{and} \quad \begin{aligned} Sg^*(1) &= \mathbb{E}_1 \exp\left(-\gamma \max_{a' \in \{0,1\}} \left\{x' - a' - \frac{\beta}{\gamma} \ln g^*(a')\right\}\right) \\ &= \exp\left(-\gamma \max\left\{2 - \frac{\beta}{\gamma} \ln g^*(0), 1 - \frac{\beta}{\gamma} \ln g^*(1)\right\}\right) \\ &= \exp[-\gamma/(1-\beta)] = g^*(1), \end{aligned}$$

we know that $Sg^* = g^*$. On the other hand,

$$\begin{aligned} Tv^*(1) &= \max\left\{1 - \frac{\beta}{\gamma} \ln \mathbb{E}_0 \exp[-\gamma v^*(x')], -\frac{\beta}{\gamma} \ln \mathbb{E}_1 \exp[-\gamma v^*(x')]\right\} \\ &= \max\{1 + \beta v^*(1), \beta v^*(2)\} = (2\beta - \beta^2)/(1-\beta) \neq v^*(1). \end{aligned}$$

Hence, $Tv^* \neq v^*$ and the value function is not a fixed point of the Bellman operator.

References

Abbring, J. H., J. R. Campbell, J. Tilly, and N. Yang (2018): "Very Simple Markov-Perfect Industry Dynamics: Theory," *Econometrica*, 86, 721–735.

Aguirregabiria, V. and P. Mira (2002): "Swapping the nested fixed point algorithm: A class of estimators for discrete Markov decision models," *Econometrica*, 70, 1519–1543.

Athreya, K. B. (2008): "Default, insurance, and debt over the life-cycle," *Journal of Monetary Economics*, 55, 752–774.

Bäuerle, N. and A. Jaśkiewicz (2018): "Stochastic optimal growth model with risk sensitive preferences," *Journal of Economic Theory*, 173, 181–200.

Benhabib, J., A. Bisin, and S. Zhu (2015): "The wealth distribution in Bewley economies with capital income risk," *Journal of Economic Theory*, 159, 489–515.

Bergemann, D. and J. Valimaki (2017): "Dynamic Mechanism Design: An Introduction," Tech. rep., SSRN.

Bertsekas, D. P. (2013): *Abstract dynamic programming*, Athena Scientific.

——— (2017): *Dynamic programming and optimal control*, vol. 4, Athena scientific Belmont, MA.

Bloise, G. and Y. Vailakis (2018): "Convex dynamic programming with (bounded) recursive utility," *Journal of Economic Theory*, 173, 118–141.

Bloom, N., M. Floetotto, N. Jaimovich, I. Saporta-Eksten, and S. J. Terry (2018): "Really uncertain business cycles," *Econometrica*, 86, 1031–1065.

Duffie, D. (2010): *Dynamic Asset Pricing Theory*, Princeton University Press.

Epstein, L. G. and S. E. Zin (1989): "Risk Aversion and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework," *Econometrica*, 57, 937–969.

Fajgelbaum, P., E. Schaal, and M. Taschereau-Dumouchel (2017): "Uncertainty traps," *The Quarterly Journal of Economics*, 132, 1641—1692.

Hansen, L. P. and T. J. Sargent (2008): *Robustness*, Princeton university press.

Hansen, L. P., T. J. Sargent, G. Turmuhambetova, and N. Williams (2006): "Robust control and model misspecification," *Journal of Economic Theory*, 128, 45–90.

Higashi, Y., K. Hyogo, and N. Takeoka (2009): "Subjective random discounting and intertemporal choice," *Journal of Economic Theory*, 144, 1015–1053.

Hong, H. and M. Shum (2010): "Pairwise-difference estimation of a dynamic optimization model," *The Review of Economic Studies*, 77, 273–304.

Jovanovic, B. (1982): "Selection and the evolution of industry," *Econometrica*, 649–670.

Ju, N. and J. Miao (2012): "Ambiguity, learning, and asset returns," *Econometrica*, 80, 559–591.

Kaplan, G. and G. L. Violante (2017): "A model of the consumption response to fiscal stimulus payments," *Econometrica*, 85, 693–734.

Kochenderfer, M. J. (2015): *Decision making under uncertainty: theory and application*, MIT press.

Kristensen, D., P. Mogensen, J. M. Moon, and B. Schjerning (2018): "Solving Dynamic Discrete Choice Models Using Smoothing and Sieve Methods," Tech. rep., University of Copenhagen.

Krusell, P. and A. A. Smith, Jr (1998): "Income and wealth heterogeneity in the macroeconomy," *Journal of Political Economy*, 106, 867–896.

Livshits, I., J. MacGee, and M. Tertilt (2007): "Consumer bankruptcy: A fresh start," *American Economic Review*, 97, 402–418.

Ljungqvist, L. and T. J. Sargent (2018): *Recursive Macroeconomic Theory*, MIT Press, 4 ed.

Marinacci, M. and L. Montrucchio (2010): "Unique solutions for stochastic recursive utilities," *Journal of Economic Theory*, 145, 1776–1804.

McCall, J. J. (1970): "Economics of information and job search," *The Quarterly Journal of Economics*, 113–126.

Norets, A. (2010): "Continuity and differentiability of expected value functions in dynamic discrete choice models," *Quantitative economics*, 1, 305–322.

Oh, H. and R. Reis (2012): "Targeted transfers and the fiscal response to the great recession," *Journal of Monetary Economics*, 59, S50–S64.

Pashchenko, S. (2013): "Accounting for non-annuitization," *Journal of Public Economics*, 98, 53–67.

Pissarides, C. A. (2000): *Equilibrium Unemployment Theory*, MIT press.

Rust, J. (1987): "Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher," *Econometrica*, 999–1033.

——— (1994): "Structural estimation of Markov decision processes," *Handbook of econometrics*, 4, 3081–3143.

——— (1996a): "Estimation of dynamic structural models, problems and prospects: discrete decision processes," in *Advances in econometrics: Sixth world congress*, Cambridge Univ Pr, vol. 2, 119–170.

——— (1996b): "Numerical dynamic programming in economics," *Handbook of computational economics*, 1, 619–729.

Schaal, E. (2017): "Uncertainty and unemployment," *Econometrica*, 85, 1675–1721.

Schorfheide, F., D. Song, and A. Yaron (2018): "Identifying long-run risks: A Bayesian mixed-frequency approach," *Econometrica*, in press.

Stokey, N., R. Lucas, and E. Prescott (1989): *Recursive Methods in Economic Dynamics*, Harvard University Press.

Tallarini, T. D. (2000): "Risk-sensitive real business cycles," *Journal of Monetary Economics*, 45, 507–532.

Wang, C., N. Wang, and J. Yang (2016): "Optimal consumption and savings with stochastic income and recursive utility," *Journal of Economic Theory*, 165, 292–331.

Winberry, T. (2018): "A Method for Solving and Estimating Heterogeneous Agent Macro Models," Tech. rep., University of Chicago.